AD-A250 867

Research Product 92-03

Methods for Determining Resource and Proficiency Tradeoffs Among Alternative Tank Gunnery Training Methods

April 1992

Fort Knox Field Unit
Training Systems Research Division

U.S. Army Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

---

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

Scott E. Graham
Joseph D. Hagman

NOTICES

FINAL DISPOSITION: This Research Product may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: This Research Product is not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992, April	Final Report Feb 90 - Sep 91

4. TITLE AND SUBTITLE
Methods for Determining Resource and Proficiency Tradeoffs Among Alternative Tank Gunnery Training Methods

5. FUNDING NUMBERS
DAHC-89-C-0009
62785A
790
3205
C1

6. AUTHOR(S)
Hoffman, R. Gene, and Morrison, John E. (HumRRO)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Human Resources Research Organization
66 Canal Center Plaza
Suite 400
Alexandria, VA 22314

8. PERFORMING ORGANIZATION REPORT NUMBER
--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
Fort Knox Field Unit
ATTN: PERI-IK
Fort Knox, KY 40121-5620

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Product 92-03

11. SUPPLEMENTARY NOTES
Contracting Officer's Representative, Barbara A. Black. A companion report (Morrison & Hoffman, 1981) provides a user's introduction to the research method discussed in this report.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT (*Maximum 200 words*)

Tank Gunnery training devices are designed to decrease costs and other resources required for training. To realize resource savings and, at the same time, maintain desired proficiency, cost-effective tradeoffs must be made between device training and on-tank training. To determine tradeoff specifications, experimental and non-experimental methods were reviewed, critiqued, and synthesized into a set of recommendations. A surrogate method, simulated transfer, which uses judgments from subject matter experts, was modified for gunnery training problems. Also delineated were nonlinear models needed to guide analysis of learning and tradeoff data.

Researchers identified an Army National Guard setting for testing the nonexperimental and simulated transfer research methods. Research focused on the Guard Unit Army Device Full-Crew Interactive Simulation Trainer, Armor (GUARD FIST I). Proficiency ratings, collected on all gunnery training events, showed some ability to predict relative performance from one training event to another, but mean ratings varied dramatically between events, suggesting that the interpretation of the rating scale shifted between raters and events. Amount of practice was negatively

(Continued)

14. SUBJECT TERMS
Training devices
Training management
Military training
Gunnery trainers
Tanks (combat vehicles)

15. NUMBER OF PAGES
109

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

13. ABSTRACT (Continued)

correlated with Table VIII, apparently because of a training strategy that gave priority in training time to crews having lower proficiency.

The data generated by the simulated transfer methods were reliable but did not fit minimal expectations. For example, the amount of training required to achieve a minimum level of proficiency was severely underestimated compared with the training that actually occurred during the period of the research.

Observations of the case-study approach offer suggestions for improving training processes. For example, researchers found that GUARD FIST I offers a significant time-savings advantage over tank-table training but that instructional guidance for this device does not exist.

Research Product 92-03

Methods for Determining Resource and Proficiency Tradeoffs Among Alternative Tank Gunnery Training Methods

R. Gene Hoffman and John E. Morrison
Human Resources Research Organization

Field Unit at Fort Knox, Kentucky
Barbara A. Black, Chief

Training Systems Research Division
Jack H. Hiller, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

April 1992

Army Project Number
2Q162785A790

Human Performance Effectiveness and Simulation

Approved for public release; distribution is unlimited.

FOREWORD

In a time of decreasing budgets and increasing training costs, the Army has adopted simulation as a cost-effective alternative to field training. Evaluating the relative effectiveness of different gunnery training methods (e.g., training devices, dry-fire gunnery, and live-fire gunnery) is a complex issue, involving tradeoffs between the resources available for training, gunnery, and the effect of those resources on gunnery proficiency. Research presented in this report is part of a project to develop research methods to determine those tradeoffs. This report provides a technical discussion of the methods for designing the research and analyzing the data. These methods represent an important contribution to the Exploratory Development program of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and will facilitate execution of well-directed tradeoff research for tank gunnery training.

This research is part of the ARI task entitled "Application of Technology to Meet Armor Skills Training Needs." It was performed under the auspices of ARI's Armor Research and Development Activity at Fort Knox. The proponent for the research is the Deputy Chief of Staff, Training, of the Training and Doctrine Command. The requirement for this research has also been recognized by the Office, Secretary of Defense.

EDGAR M. JOHNSON
Technical Director

METHODS FOR DETERMINING RESOURCE AND PROFICIENCY TRADEOFFS AMONG ALTERNATIVE TANK GUNNERY TRAINING METHODS

EXECUTIVE SUMMARY

Requirement:

Tank gunnery training devices are designed to decrease the costs and other resources required for training, but they are suspected of deficient training capabilities compared with live-fire training on the tank. To realize the savings of training devices and, at the same time, maintain desired proficiency, cost-effective tradeoffs must be made between investments in device training and on-tank training. Research methods are required to generate these tradeoff specifications. This report presents various research methods that address this tradeoff problem. It also presents tests for some of those methods on an operational training problem in tank gunnery.

Procedure:

The requirement to specify research methods was treated as four sub-problems. First, experimental methods with controlled manipulations were reviewed, critiqued, and synthesized into a set of recommendations. Second, because well-controlled experimental research is so difficult to implement in an operational training context, nonexperimental methods were examined as an alternative. Third, various nonlinear models were considered to guide analysis of data by these research methods. Fourth, a surrogate method of "simulated transfer" research was modified for tank gunnery training problems. This method uses judgments from subject matter experts as an alternative to empirical performance data.

Researchers identified an Army National Guard setting for testing the nonexperimental and simulated transfer research methods. Research focused on the Guard Unit Armory Device Full-Crew Interactive Simulation Trainer, Armor (GUARD FIST I), a tank-appended computer-based training system. Crews were trained for Table VIII by using a variety of gunnery training events (e.g., Tables I-VIII) and training methods (e.g., GUARD FIST I, dry-fire, and live-fire). Although the events were not under experimental control, performance measurement was standardized by using a performance rating form developed to track proficiency across different gunnery training events.

Findings:

Several options with varying degrees of sophistication were specified for experimental, nonexperimental, and simulated transfer methods. The methods all require the measurement or estimate of proficiency at multiple points during training. Further, in contrast to training transfer research, training and multiple assessments of proficiency must be considered on the operational equipment as well as on the alternative training devices.

Data analysis requirements were also presented. Because the resource-proficiency relationships under investigation are expected to be curvilinear, nonlinear analysis methods are required. Nonlinear methods require equations that can relate either training time or cost to expected proficiency. Numerous learning and tradeoff functions were presented that differed in form (e.g., power function, exponential functions), parameters (e.g., rate of learning, previous experience, cost ratios), number of training methods included (one, two, more than two), and method of combining training alternatives (additive or interactive). Solving the tradeoffs between device and operational equipment training requires estimating parameters for these equations. Although nonlinear analysis methods must be used, they are not straightforward and interpretable solutions are not guaranteed.

The nonexperimental method (a correlational, case-study approach) was unsuccessful in tracking performance improvements over successive training events. Proficiency ratings showed some ability to predict relative performance from one training event to another, but mean ratings varied dramatically between events, suggesting that the interpretation of the rating scale shifted between raters and events. Also, amount of practice was negatively correlated with performance on Table VIII, apparently because of a training strategy that gave priority in training time to crews having lower proficiency.

The simulated transfer methodology was also unsuccessful. The data generated were reliable, but they did not fit minimal expectations. For example, the amount of training required to achieve a minimum level of proficiency was severely underestimated compared with the training that actually occurred during the period of the research.

Observations of the case-study approach were used to make suggestions for improving training processes. For example, GUARD FIST I offers a significant time savings advantage over tank-table training. A six- to tenfold increase in the number of engagements practiced is possible using GUARD FIST I. However, guidance in instructional techniques for the device does not exist. A GUARD FIST I training manual should be developed.

Utilization of Findings:

This research found no easy answers to the problem of determining resource tradeoffs among alternative training methods. Data collection and analysis requirements are complex. To address tradeoff questions empirically, significant research resources must be committed to the effort and, even then, the focus may need to be kept fairly narrow. Policy makers and research planners have typically underestimated the effort needed to address training tradeoff questions. Simple case-study observations by outside researchers can provide valuable technical assistance in implementing and improving training procedures, but they provide little information on resource tradeoffs.

METHODS FOR DETERMINING RESOURCE AND PROFICIENCY TRADEOFFS AMONG ALTERNATIVE TANK GUNNERY TRAINING METHODS

CONTENTS

CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

CONTENTS (Continued)

CONTENTS (Continued)

Page

CONTENTS (Continued)

METHODS FOR DETERMINING RESOURCE AND PROFICIENCY TRADEOFFS AMONG ALTERNATIVE TANK GUNNERY TRAINING METHODS

Chapter 1. Introduction and Background

Problem

Armor decision makers often turn to researchers to answer questions about the effectiveness of gunnery training devices. An often-asked and fundamental question is whether or not training devices improve gunnery performance on the tank. Researchers typically propose to answer this question by performing a transfer-of-training experiment. In this sort of experiment, the gunnery performance of a group of individuals or crews that have received previous training on a device is compared to the gunnery performance of another group that has not received such pretraining. Transfer experiments such as these have been performed on specific gunnery training devices such as the Unit Conduct-of-Fire Trainer (U-COFT), the Videodisk Integrated Gunnery Simulator (VIGS), and the arcade-like TopGun training device (Morrison, Drucker, & Campshure, 1990).

The basic transfer-of-training experiment answers an important question: "Does a training device have an effect on job performance?" However, the decision makers who procure, evaluate, and use training devices are now starting to ask more difficult questions, such as "how much is a pound of training worth?" Although not stated in precise and testable terms, this question implies an interest in the monetary bottom line for training. In that same vein, more testable questions can be formulated:

- What are the gains in proficiency that are accrued as a function of training costs?
- Do the transfer benefits of new devices justify their operational costs?
- To what extent can practice on the device be effectively substituted for practice on the actual equipment?
- What is the optimal mix of training on device(s) and actual equipment?

These more complex questions can only be answered by examining the relationships and tradeoffs between *training resources* and *performance proficiency*. Training resources refer to expenses that are incurred in the process of training. Resources are quantified by measures such as the number of practice repetitions (or trials), the time of training, or the cost of training. Proficiency is defined in terms of the accumulation of knowledges and skills required to successfully perform the job. This variable is usually measured in terms of individual-, crew-, or platoon-level performance on the device itself, on another related training device, or on the actual equipment (the tank). Thus, for a given training method, a training resource-proficiency *relationship* is described by the learning curve for that method. Obviously, alternative training methods may exhibit different relationships between resources and proficiency. A training resource-proficiency *tradeoff* refers to the more complex comparison of two or more resource-proficiency relationships. Therein lies the nature of the tradeoff: the substitutability

of one training method for another in terms of both resulting proficiency and costs.

Research in the Present Series

The present research is part of a series of reports that concern the development of research methods and tools for investigating resource-proficiency tradeoffs. The project began by developing research methods for determining threat scenarios for tank gunnery (R. C. Campbell & C. H. Campbell, 1990). Doyle (1990) applied these research methods and derived 42 target scenarios and 7 scenario enhancements. The detailed scenarios included information such as the number, type, and range of likely threat vehicles as well as their formation and rate of movement. C. H. Campbell and Hoffman (1990) elaborated on this research by developing methods for selecting threat scenarios based on the training requirements of actual units. The purpose of this initial research was to define realistic combat conditions for training gunnery.

Another aspect of the present research was the development of training that was appropriate to the threat-based environment. The training component of the project was addressed in two stages. First, research methods were developed for determining training objectives that were appropriate to the threat scenarios (Morrison, Meade, & R. C. Campbell, 1990). Objectives were defined as subtasks performed at the crew or platoon level. Analysis indicated that these subtasks formed an extensive and heterogeneous set of training objectives. Second, strategies were examined for training these gunnery objectives (Morrison & Holding, 1990). Training strategies were defined as research methods for solving general problems in training design. These strategies were discussed in the context of four problems: (a) determining the appropriate structure of training objectives, (b) determining the appropriate sequencing of training objectives, (c) selecting appropriate devices for training the objectives, and (d) allocating training time to each objective/device combination. Training strategies were proposed for some typical armor training problems.

Hoffman, Fotouhi, Meade, and Blacksten (1990) addressed the performance measurement aspect of the project. Analysis of gunnery performance indicated that performance should be measured in terms of both performance outcomes (i.e., target hits and misses) and behavioral processes. For measuring the outcome of performance, these researchers advocated using a composite measure, the hit expectation ratio, which was defined as the expected number of Blue hits on Red divided by the expected number of Red hits on Blue. This measure currently underlies the scoring for Table VIII (the crew gunnery qualification exercise) that comprises 10 gunnery engagements having either one or two targets. Hoffman et al. extended the hit expectation metric to score threat target arrays of up to five targets. For measuring behavioral processes, the researchers devised a system for rating behaviors on evaluation criteria for the subtasks devised by Morrison, Meade, and R. C. Campbell (1990).

Previous research in this series has provided the necessary background for performing research on the resource-proficiency tradeoffs that pertain to tank gunnery training. This report presents an integration of these research methods and tools by suggesting how they might be used to answer specific questions about the relationship between training resources and performance proficiency.

Organization of the Present Report

Chapters 2-4 of the present report describe some of the important research methods and tools that are specifically related to the analysis of resource-proficiency relationships and tradeoffs. Chapter 2 presents both experimental and nonexperimental designs that are appropriate for determining these relationships and tradeoffs. Chapter 3 presents data analysis techniques that are unique to these problems. Chapter 4 discusses the problems in obtaining performance-based data for this research and presents alternative methods for obtaining surrogate measures of performance. At the conclusion of these chapters, the major points are reiterated in the form of recommendations intended for researchers who are planning to conduct tradeoff research in the context of tank gunnery.

Chapters 4 and 5 present tests of selected research methods to determine their validity and usefulness. The latter half of Chapter 4 continues the discussion of alternative data techniques by providing a test of simulated transfer methods using estimates provided by soldiers in a Reserve Component unit. Chapter 5 returns to the topic of performance-based research methods and presents a test of a nonexperimental design method: the case-study technique.

A companion research report (Morrison and Hoffman, 1991) is available which provides a less technical summary of the basic concepts elaborated in this report. This report is a detailed, technical presentation for research concepts and strategies for addressing trade-off among training resources and training proficiency.

Chapter 2. Research Designs

The present chapter describes research designs that are appropriate for determining resource-proficiency relationships and tradeoffs. For our purposes, research designs are defined as plans for collecting data. These plans outline (usually in schematic form) the order of treatments or treatment combinations administered to each experimental condition. Not included in the present chapter is a discussion of experimental control issues that pertain to applied experimentation; authoritative discussions of these issues are provided elsewhere (e.g., Campbell & Stanley, 1966; Cook, Campbell, & Peracchio, 1990; Hagin, Osborne, Hockenberger, Smith, & Gray, 1982; Boldovici, 1987).

In addition to the classic transfer-of-training paradigm described in Chapter 1, there are a wide variety of other research designs that may be appropriate for investigating the resource-proficiency relationships and tradeoffs. To impose order on this heterogeneous domain, the designs described in the present chapter are divided into two broad categories: experimental and nonexperimental. Within each of those two design categories, the chapter compares and contrasts specific prototypical designs. Where appropriate, examples are provided of the designs used in previous armor research or in related areas of military training research.

Experimental Designs

Experimental designs are those wherein the researcher exerts direct control over independent variables (i.e., training resources) and measures their effects on dependent variables (performance). The resulting data can be used to determine cause and effect relationships between training resources and performance proficiency. The following experimental designs are divided into three subcategories that are presented in order of increasing complexity and sophistication.

Performance Improvement on a Device

The two designs in this subcategory address the changes to performance on a training device that occur as a function of practice. The essence of these "pretransfer" designs is that a single experimental group is repeatedly trained and tested on a device.

Two-point assessment. The top row of Figure 1 illustrates the minimal requirements of this design: that is, that performance of a single group (G1) is assessed immediately before and after training on Device A. The performance assessment is executed on the device itself or off line, using an instrument such as a paper-and-pencil test that is designed to assess skills and knowledges learned on the device.[1] The bottom row illustrates an often-used elaboration on this design where the performance of an experimental group (G1) is compared to a control group (G0) that receives the pre- and posttests but does not receive training on Device A. The purpose of the control group

---

[1]If the pre- and posttests are assessed on the actual equipment, this design could be used to measure transfer of training, rather than improvement in performance on the device.

Group		Events		
		M	T	M
G1		A	A	A
G0		A	-	A

Figure 1. Two-point design for assessing performance improvement on a device. (T indicates a training event and M, a measurement event.)

is to separate the effects of device training per se from the effects of repeated testing on the device.

Kraemer and Smith (1990, Experiment 3) used such a two-group design to determine the effects of TopGun practice. Both their experimental and control groups received pretests and posttests administered on the TopGun device. In addition, the experimental group received a period of practice on TopGun between the pre- and posttests. Results from this experiment showed reliably greater increases in performance from pre- to posttest for the experimental group compared to the control group. Thus, this experiment demonstrated that learning occurs as a function of practice on TopGun.

Multiple-point assessment. A limitation of the two-point design is that it does not allow the researcher to determine the functional relationship between training resources (i.e., time and costs) and performance on the device; to do so, requires the measurement of performance at more than two points. The appropriate number of points, in turn, depends on the expected form of that function. As illustrated in Figure 2, researchers commonly assume that performance is an increasing, negatively accelerated function of practice trials that approaches an asymptotic point indicating a ceiling to performance. (Mathematical formulations of this relationship are discussed in Chapter 3.) Estimation of the curvilinear portion of the function requires a minimum of 3 data points and its asymptote requires an additional fourth point.

The resulting minimum requirements for this design are outlined in Figure 3, which indicates that a single group (G4) is trained and tested repeatedly over 4 blocks of trials. This figure also illustrates the case in which training and measurement trials are integrated; that is, performance is measured on the repeated individual or blocks of practice trials.

An example of the multiple-point assessment is provided by an experiment that examined the effects of repeated practice on the TopGun gunnery training device (Bliss, 1989). In this experiment, the performance of 20 college students was measured over four blocks of TopGun gunnery engagements, with each block consisting of 36 engagements. Examination of means and variances showed that the students rapidly improved on TopGun with their performance stabilizing by the second block of engagements.

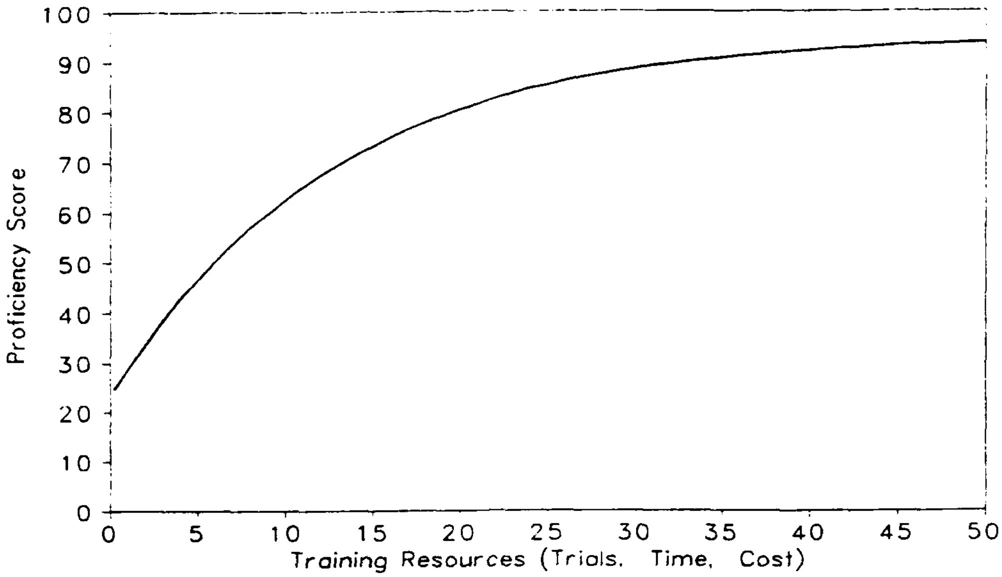

Figure 2. Example learning curve describing proficiency on a training device as a function of resources devoted to training on the device. (Scales and function are hypothetical.)

	Events			
	T/M	T/M	T/M	T/M
G4	A	A	A	A

Figure 3. Multiple-point design for assessing performance improvement on device A. Note that training (T) and measurement (M) are integrated into a single event (T/M).

Transfer-of-Training Designs

The transfer-of-training design addresses a fundamental relationship between performance on two training methods: whether or not skills learned on training Method A transfer to another training Method B. Note that Method B may represent training on the actual equipment or on another training device.

Two-group design. As illustrated in Figure 4, the classic transfer design stipulates that two groups be compared: an experimental group that receives training on a Device A and a control that does not receive this training. The two groups then receive a performance test on training Method B. Any performance differences between the groups on B can be attributed to transfer of training from Device A. Elaborations on this basic transfer design are discussed by Postman (1971).

For most gunnery research situations, it would be unsafe to have soldiers in the control condition operate a weapons system without some prior

		Events	
		T	M
Group	G1	A	B
	G0	-	B

Figure 4. Basic two-group design for determining transfer of training from Device A to performance on B, which may represent the actual equipment or another device. (T indicates a training event and M, a measurement event.)

training. An alternative is to train the control group but on a different training method (C). For instance, a control group may be assigned to a group that receives a lecture on device operation. Note, however, that the transfer question is no longer whether training on Device A transfers to B; rather, the issue becomes whether the training transfer from A to B is more or less than the transfer from C to B.

An example of the relative transfer design is provided by an experiment performed to evaluate VIGS (Boldovici, 1986). In this experiment, a control group that received conventional (non-device) training on gunnery was compared to two experimental groups that received the same conventional training plus training on one of two different devices. One experimental group received training on VIGS and the other received training on an older mechanical training device, the Wiley Burst-on-Target trainer. All three groups were tested using a dry-fire exercise on the actual tank. That test indicated faster opening times and fewer procedural errors for the VIGS group compared to the other two groups. In other words, Boldovici's data suggested that VIGS was more effective than the Wiley device in supplementing conventional training.

Multi-group design. The classic transfer-of-training design has been criticized for not allowing the measurement of the effects of amount of device training (Boldovici, 1987). As shown in Figure 5, a simple modification of the design would be to employ multiple (i.e., 3 or more) experimental groups having different amounts of training on Method A before transferring to training Method B. Results from this sort of design would enable the researcher to fit a curve describing performance on B as a function of training on A.[2] This design has the added advantage of allowing the researcher to determine the effects of practice on A itself. That is, a learning curve for Method A can be derived from the groups that receive repeated blocks of practice on Method A (Groups G2 and G3 in Figure 5).

[2]The levels of A training need not be evenly spaced as shown in the figure. In fact, because of the negatively accelerated shape of the typical transfer function, there is an advantage to having more points at the beginning of training before the curve begins to flatten out (see Figure 2).

Group		Events			
		T/M	T/M	T/M	M
	G3	A	A	A	B
	G2	-	A	A	B
	G1	-	-	A	B
	G0	-	-	-	B

Figure 5. Multi-group transfer-of-training design for determining transfer from Device A to performance on B, which may represent the actual equipment or another device. (T indicates a training event and M, a measurement event.)

Hart, Hagman, and Bowne (1990) used a design similar to that shown in Figure 5 to measure the effects of transfer from the TopGun part-task trainer to the Mobile Conduct-of-Fire Trainer (M-COFT) full-task gunnery simulator. They compared the M-COFT performance of two experimental groups that were pretrained on TopGun to the M-COFT performance of a control group that was not pretrained on TopGun. One experimental group received three 20-min blocks of TopGun training whereas the other received only one 20-min block of TopGun training. The two experimental groups that received TopGun pretraining were significantly more accurate than the control group in hitting stationary targets using the gunner's auxiliary sight. However, there were no differences between the two experimental groups. Although the researchers did not have enough data points to fit a transfer function, the results suggested that transfer from TopGun to M-COFT was rapid and virtually complete after only a single 20-min TopGun training session.

Tradeoff Designs

The previous transfer designs assess the effect of pretraining on Method A on **initial** performance on Method B. Hammerton (1989) has described these as "first shot" measures of transfer and cautioned that they may not reveal the ultimate value of training on A. For example, he cited research showing negative first-shot transfer of a training condition that eventually resulted in a 70 percent savings in B training. Spears (1985) has argued that this phenomenon is due to the effects of "interfacing," which is defined as the interference of device-related actions learned in A that must be adapted to the stimulus context of B. The effects of interfacing are a short-lived deficit in performance and are quickly overcome by the positive effects of the skills learned on A. To assess the potential transfer savings on B as a result of training with A, the researcher must plan to train on B as well as on A. Furthermore, this is a requirement for determining the tradeoff between these two training resources. That is, even without the interfacing problem, making a resource-proficiency tradeoff comparison requires resource-proficiency relationships for both training methods under consideration. Thus, the defining characteristic of tradeoff designs is that the trainee

receive repeated practice on the transfer (B) method as well as on A. Two variations on this type of design are described below.

Roscoe design. The first sort of design is that which was originally described by Roscoe (1971, 1972) as a hypothetical experiment for determining cost-effectiveness of training devices. In his design, groups are differentiated on the amount of training that they receive on A and are then trained to a common performance criterion (i.e., standard) on B. In Figure 6, this is shown by repeated training on B for an indefinite number of trials. The dependent variable in this design is the amount of training on B required to reach the criterion.

		Events			
		T/M	T/M	T/M	T/M ....
Group	G3	A	A	A	B ....
	G2	-	A	A	B ....
	G1	-	-	A	B ....
	G0	-	-	-	B ....

Figure 6. Roscoe design for determining transfer savings on B as a function of amount of training on A. The integrated training and measurement (T/M) events on B continue for an indefinite number of trials until a performance standard is met.

There are no examples of this design being used to evaluate gunnery training devices; however, the Roscoe design has been used in two experiments designed to evaluate aviation training devices (Povenmire & Roscoe, 1973; Bickley, 1980). An important innovation of Bickley's (1980) research was his derivation of a research method for determining the mix of device and aircraft training that minimized training costs. The method was based on a function that relates the amount of subsequent training required to reach criterion on the actual equipment (Method B) to the amount of training on the training device (Method A). As illustrated in Figure 7, this function is sometimes referred to as an "iso-performance" function because each point on the curve specifies a mix of trials on the device and trials on the actual equipment that result in a set standard of performance. At about the same time, Carter and Trollip (1980) derived a research method using similar iso-performance functions to determine the mix of device and aircraft training that maximizes performance for a given amount of training time or costs. This points to one key advantage to the Roscoe design: it permits the researcher to derive a tradeoff function of two training methods under certain constraints (i.e., fixed performance or fixed costs). The tradeoff functions, in turn, can be used to determine optimal allocations of training resources.

Groups-by-trials design. An alternative to training on B to a fixed performance standard is to train on B for a fixed number of trials. As

illustrated in Figure 8, this design manipulates two independent variables (amount of training on A and amount of training on B) with performance on the

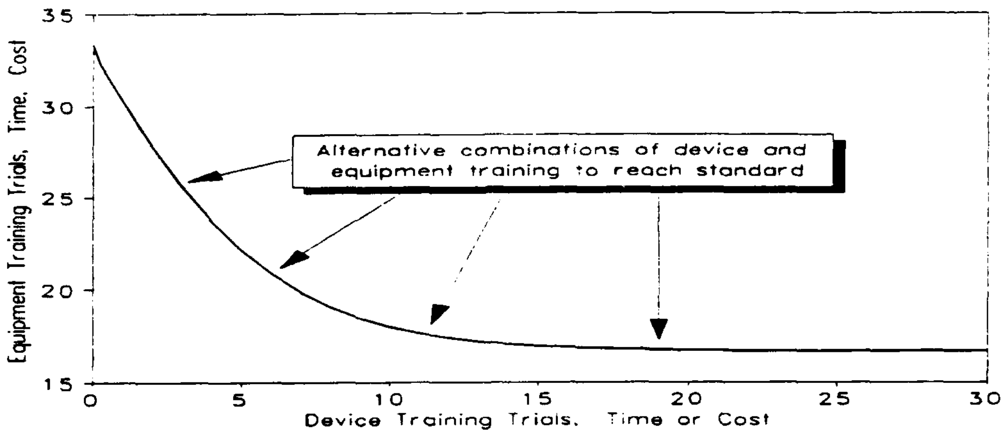

Figure 7. Example iso-performance function.

		Events						
		T/M	T/M	T/M	T/M	T/M	T/M	T/M
Group	G3	A	A	A	B	B	B	B
	G2	-	A	A	B	B	B	B
	G1	-	-	A	B	B	B	B
	G0	-	-	-	B	B	B	B

Figure 8. Groups-by-trials design for examining the interaction of amount of A training on repeated trials of B training. As indicated, the integrated training and measurement (T/M) events occur for a fixed number of trials.

Method B providing the dependent variable. The advantage to this design is that it allows the researcher to determine whether the amount of training on A changes the shape (i.e., rate, asymptote) of the acquisition curve for B or simply changes the starting point on a single learning curve. Although this design would provide basic information for allocating training time between two training methods, it has not been used in the context of research on armor gunnery.

The lack of armor examples of tradeoff designs is partly due to the fact that they may not be feasible for many gunnery training situations. For

instance, a training manager may want to know the relationship between training on a computer-based gunnery training device (e.g., U-COFT) and performance on a live-fire gunnery exercise (e.g., Table VIII). Live-fire exercises are not repeatedly administered to individual crews for many reasons including high costs of ammunition and fuel, constraints on live-fire ranges, and the time required to administer the exercise. Because of the manner in which Table VIII is administered, the one-shot transfer design may be the only alternative for the researcher. It should be further cautioned that the one-shot design eliminates the possibility of examining an issue of much interest to gunnery trainers--that is, the substitutability of device training for live-fire training. Tradeoff issues can only be addressed when repeated trials occur on each of the alternative training methods.

Extensions Beyond Two Training Methods

Up to this point, we have presented designs for examining resource-proficiency relationships within a single device or the transfer/tradeoff from a device to another method for training or performance testing. In this section, we examine extensions of these ideas to situations where there are three or more training methods under consideration. It is assumed that the optimal training order among the various training methods has either been determined for the sake of simplicity.

The three-methods case. The three previous tradeoff designs (multi-group transfer, Roscoe, and groups-by-trials) can be extended to examine the mix of training on three training training methods: A, B, and C. As shown in Figure 9, the most straightforward extension of the design is to form 16 separate groups that receive all combinations of four levels of training on A and B. All 16 groups are treated identically with respect to training on C. The design variations differ with respect to the number of trials conducted on C. Variation 1 (a one-shot transfer to training method C) may be used to assess tradeoffs between training Methods A and B. To assess tradeoffs among all three training methods requires either variation 2 (training to a set standard on C) or variation 3 (training for a set number of trials on C).

The soldier and logistical requirements associated with managing 16 different experimental conditions is beyond the support capability of most armor units (and research organizations, for that matter). One approach to reducing the number of groups is to employ a fractional factorial design. Fractional factorial designs are constructed by systematically confounding one of the three between-group effects (A, B, or the A X B interaction). For instance, if the effects of A and B were assumed to be additive, then the researcher could systematically confound the A X B interaction, which would yield an 8-group half-fraction design. This design would enable the investigator to assess the main effects of A or B. Furthermore, because C is a repeated measures variables and is fully crossed with both A and B effects, the A X C and B X C interactions could also be assessed. However, the systematic confounding would render the A X B and the A X B X C interactions uninterpretable.

The three-methods case transfer design has not been used in tank gunnery research, although there is potential for application. For instance, one might imagine a program wherein gunner training begins with a part-task device like VIGS (A), transfers to a higher fidelity (and higher cost) method like U-COFT (B), and ends up with training on the tank itself (C). Variation 1

Group	Events						Variation*					
							1	2	3			
	T/M	T/M	T/M	T/M	T/M	T/M	T	T/M...	T/M	T/M	T/M	T/M
G33	A	A	A	B	B	B	C	C ...	C	C	C	C
G32	-	A	A	A	B	B	C	C ...	C	C	C	C
G31	-	-	A	A	A	B	C	C ...	C	C	C	C
G30	-	-	-	A	A	A	C	C ...	C	C	C	C
G23	-	A	A	B	B	B	C	C ...	C	C	C	C
G22	-	-	A	A	B	B	C	C ...	C	C	C	C
G21	-	-	-	A	A	B	C	C ...	C	C	C	C
[illegible]	-	-	-	-	A	A	C	C ...	C	C	C	C
G13	-	-	A	B	B	B	C	C ...	C	C	C	C
G12	-	-	-	A	B	B	C	C ...	C	C	C	C
G11	-	-	-	-	A	B	C	C ...	C	C	C	C
G10	-	-	-	-	-	A	C	C ...	C	C	C	C
G03	-	-	-	B	B	B	C	C ...	C	C	C	C
G02	-	-	-	-	B	B	C	C ...	C	C	C	C
G01	-	-	-	-	-	B	C	C ...	C	C	C	C
G00	-	-	-	-	-	-	C	C ...	C	C	C	C

*Variations 1, 2, and 3 represent extensions of the multi-group transfer design, the Roscoe design, and the groups-by-trials design, respectively.

Figure 9. Extensions of previous designs for measuring transfer among three media.

from Figure 9 could be used to determine the optimal allocation of training time between VIGS and U-COFT, whereas variations 2 and 3 could be used to determine the optimal allocation among all three training methods.

<u>Beyond three methods</u>. These concepts can be further extended to the 4-methods case. However, the number of separate groups becomes quickly unmanageable. Assuming that training is repeated on a fourth training method (D), there are 64 combinations of four amounts of training on the three preceding devices (A, B, and C). One approach to reducing the number of conditions is to use a half-fraction design analogous to the one discussed above, which would require 32 separate conditions. Yet even this systematically confounded design appears practically unsupportable.

Two alternative designs can provide somewhat greater economies. One alternative is a 4 X 4 latin-square design, which would require 16 different experimental conditions. This design typically confounds the triple (A X B X C) interaction, but leaves all main effects and two-factor interactions interpretable. A second alternative, suggested by Carter and Trollip (1980), is to use response surface methodology (RSM) to investigate the transfer among multiple training methods. As described by Clark and Williges (1973), RSM is an approach to experimental design and analysis that minimizes the number of conditions required to determine the functional relationships between variables. For the present case, the RSM design would require a minimum of 11

different conditions to determine the functional transfer relations among the training methods.[3]

Each of the two alternatives briefly described above could be outlined in more detail as in the previous examples. Such detailed outlines are not provided because their applicability to resource-proficiency problems is limited for at least two reasons. First and foremost is the fact that even the most efficient design, the RSM, requires a substantial number of experimental conditions. It is unlikely that a researcher would be able to obtain the support required to perform such a complex experiment. Second, as discussed in the next chapter, the data analysis for more than two devices is highly complex unless certain simplifying assumptions are made. These assumptions negate much of the potential of these designs to assess interactions among training alternatives.

Choosing an Experimental Design

Ideally, the choice of experimental design should be driven primarily by the issue that the researcher wishes to investigate. Table 1 summarizes the tradeoff issues addressed by the experimental designs previously discussed. The table indicates that as the designs become increasingly sophisticated, they address more and more of the tradeoff issues. That is, the more sophisticated design will tend to address the same issues as the less sophisticated design plus some additional ones that the less sophisticated design cannot. Clearly, the more sophisticated designs provide more research benefits in terms of the information they provide about resource-proficiency relationships and tradeoffs.

Table 2 provides an idea of the price to be paid for design sophistication. This table provides only a brief summary of the minimal requirements for research support. The table indicates that the more sophisticated designs require a greater number of experimental conditions. A greater number of conditions implies that a greater number of sampling units (soldiers, crews, platoons) are required. (For a discussion of minimum sample sizes in gunnery research, see Morrison, 1990.) A greater number of conditions is also costly to administer from the researcher's point of view and is more intrusive from the user's point of view. The table also indicates that the more sophisticated designs generally require more data in terms of repeated measurements of performance. These measurement requirements are costly in terms of staff time required to collect the data. In addition, these repeated measures are also more costly to process, reduce, analyze, and interpret.

Table 2 also illustrates the researcher's dilemma. Resource-proficiency tradeoff analyses are required to address the questions that policy-makers are asking about how to allocate their training budgets. Such analyses require

---

[3]The minimum RSM design requires 4 conditions from the half fraction of the 2^3 equivalent design; 6 conditions for assessing the "star component" of the design; and 1 condition to assess the centroid. Note that whereas the centroid represents only one condition, it is sampled twice; in other words, the centroid condition would require twice as many sampling units (soldiers, crews, platoons) as the other 10 conditions.

Table 1

Summary of Issues Addressed by or Not Addressed by Experimental Designs for Determining Resource-Proficiency Relationships

Experimental Designs	Issues Addressed by Designs								
	Does learning occur on A?	Learning curve for A	Does transfer occur from A to B?	Transfer function from A to B	Tradeoff of A & B training	Interaction of A & B training	Transfer of A & B training to C perf	Tradeoff of A, B, & C training	Interaction of A, B, & C training
Two-point assessment of improvement on a device	Yes	No	No	No	No	No	No	No	No
Multiple-point assessment of improvement on a device	Yes	Yes	No	No	No	No	No	No	No
Two-group transfer design	Yes/No[a]	No	Yes	No	No	No	No	No	No
Multi-group transfer design	Yes	Yes	Yes	Yes	No	No	No	No	No
Roscoe design	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Groups-by-trial design	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Three-methods case: Variation 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Three-methods case: Variation 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Three-methods case: Variation 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[a]Positive transfer to B implies that learning has occurred on A. However, lack of transfer to B does not necessarily imply that learning does not occur on A.

Table 2

Summary of Experimental Designs With Respect to Minimum Research Support Requirements

Experimental Designs	Number of Experimental Groups	Number of Repeated Training/Measurement Events on Training Method[a]		
		A	B	C
Two-point assessment of improvement on a device	1	2	---	---
Multiple-point assessment of improvement on a device	1	4	---	---
Two-group transfer design	2	0-1	1	---
Multi-group transfer design	4	0-3	1	---
Roscoe design	4	0-3	1+[b]	---
Groups-by-trial design	4	0-3	4	---
Three-methods case: Variation 1	16[c]	0-3	0-3	1
Three-methods case: Variation 2	16[c]	0-3	0-3	1+[b]
Three-methods case: Variation 3	16[c]	0-3	0-3	4

[a]The number of repeated measures on a training method may vary as a function of group assignment.

[b]The plus sign indicates that this design requires at least one trial. More probably there will be multiple trials to reach the performance criterion.

[c]Assumes a full factorial design. The number of groups may be reduced by using a fractional factorial design.

rather sophisticated research designs. Unfortunately, policy-makers and research planners have typically underestimated the effort needed to address training tradeoff questions. We suggest that researchers use the information provided in Tables 1 and 2 and work directly with policy makers to choose an appropriate experimental design--that is, one that satisfies the most important issues without incurring intolerable costs.

Nonexperimental Designs

As discussed in the previous section, experimental designs, especially those that address complex resource-proficiency tradeoff issues, can be costly in terms of numbers of soldiers and performance measurements that are required. In addition to the high costs of experimentation, two other characteristics of the operational environment weigh against the use of experimental techniques and in favor of nonexperimental research methods. First, there are limitations on the extent to which training resources can be experimentally manipulated in an applied setting. For example, a training manager is likely to object to providing additional training to some groups while denying training to others. Rather than imposing certain amounts of training, a nonexperimental design could permit the researcher to observe the effects of "natural" variations in training resources that can occur both within and among groups. Second, other uncontrolled training events occur during the course of normal unit training that can obscure the relationship between training resource(s) and proficiency. For instance, a researcher may wish to determine the relationship between training resources defined by U-COFT training and proficiency as measured on Table VIII. Other gunnery training exercises that the unit might conduct (e.g., Tables IV - VII) would make the U-COFT/proficiency relationship more difficult to determine. Observational data collection methods, coupled with correlational data analysis techniques, enable the researcher either to control for these intervening events or to examine their effects within a more comprehensive, multivariate model of gunnery proficiency.

At the outset, we should emphasize that we do not regard nonexperimental designs as the preferred approach for determining resource-proficiency relationships. The principle problem with using nonexperimental designs is the difficulty with establishing cause-effect relationships between training resources and proficiency. The essence of this problem is that the existence of covariation between training resources and proficiency does not necessarily imply that proficiency is caused or controlled by the training resources. For instance, suppose a researcher finds a positive correlation between amount of device training and proficiency. On the surface, this finding would seem to suggest that increasing the amount of training on the device increases proficiency. Another possible interpretation, however, is that the relationship between training resources and proficiency is spurious; that is, the empirical relationship can be explained by the possibility that both variables covary with a third variable. To continue with the example, suppose training on the device were available to all who wanted to use it. In that case, amount of training might then covary with the trainee's motivation to seek out training experiences. It might be possible then that the observed positive relationship between amount of training and proficiency would say more about the trainees than it does about device effectiveness. If, on the other hand, the researcher had experimentally manipulated the amount of training, then the resources variable would have been orthogonal to (i.e., uncorrelated with) all other determinants of proficiency, measured or

unmeasured. The resulting relationship between training resources and proficiency would be therefore less ambiguous.

Although not the preferred approach, the correlational method may provide a useful expedient for deriving resource-proficiency tradeoffs in those situations where experimental manipulation is out of the question. In reality, aspects of experimental and nonexperimental approaches can be incorporated effectively into a single hybrid design. To simplify exposition, however, the following discussion distinguishes among purely nonexperimental designs. These nonexperimental designs are divided into two subcategories: (a) those "pure" correlational designs wherein the researcher exerts no control over the training resources--that is, they are left free to vary; and (b) those designs wherein the researcher exerts some level of indirect control over the training resources.

Pure Correlational Designs

The advantage of pure correlational designs is that the researcher does not have to intrude in training to control the amount and type of training that soldiers receive. However, this research method does require the researcher to obtain quantitative indexes of training resources such as the amount of time devoted to each objective. These can usually be obtained through simple observation of training or other unobtrusive means such as obtaining information from training documents. The following sections describe two variations of the pure correlational design, which differ only in terms of the size of the sample.

Case-study design. In the most constrained situation, the researcher may be able only to observe a small sample of individuals, crews, or platoons that are undergoing training. Under these conditions, statistical analyses are not possible and the researcher's only alternative is to intensively examine the training process on a case-by-case basis. Although this represents the "worst case" scenario for the researcher, some valuable information can be gained from this sort of examination. In particular, the case-study approach enables the researcher to examine a training resource in a realistic situation without ignoring the myriad of other contextual factors that normally affect gunnery proficiency. Furthermore, with fewer cases, it is possible to collect more detailed quantitative and qualitative data than would be possible with a large-sample investigation.

The case study can address a variety of questions about training. One question that is relevant to resource-proficiency tradeoffs is whether or not trainees can perform the tasks in question on the operational equipment after receiving training on a device. For instance, flight training often begins with device-based training on basic flying skills and procedures prior to the first flight in the aircraft. The fact that the first flight is usually successful to some degree is regarded as evidence in favor of the device (Caro, 1977). A clear problem with this reasoning is that operational performance obtained from this design cannot be attributed to device training alone. To use an armor training example, suppose that a researcher demonstrated that entry-level soldiers successfully hit actual targets after training with a device. The problem is that successful performance might be due to any gunnery training that soldiers normally receive (e.g., classroom instruction, on-tank exercises) in addition to training on the device. It

would not be unreasonable to suppose that the additional training is at least partly responsible for the performance of soldiers.

In addition to this problem of inference, another disadvantage of the case study approach is that it provides only minimal information relevant to resource-proficiency relationships. At best, any quantitative data gathered on small samples should be treated as only suggestive of actual relationships. For instance, Kraemer and Bessemer (1987) examined the effect of training with the Simulation Networking (SIMNET) system on the performance of the U.S. team for the Canadian Army Trophy (CAT) competition. CAT performance was scored at the platoon level, with only 9 U.S. M1 platoons participating. In addition to examining SIMNET training practices in detail, they calculated the correlation between the number of SIMNET battleruns completed and the score obtained on the CAT competition. They found a moderate positive correlation ($r = .53$) between amount of training and performance, suggesting that variations in amount of SIMNET training had an effect CAT performance. With only 7 degrees of freedom, however, the correlation was not significant using traditional standards for statistical reliability. Furthermore, authors were quick to point out that the correlation was mostly accounted for by 1 platoon that completed more SIMNET exercises than the other 8 and also scored high on CAT. In short, the data gathered by Kraemer and Bessemer were not sufficient to derive a resource-proficiency relationship.

Large-sample correlational design. Given larger samples, the data obtained from correlational research can be treated with more sophisticated statistical techniques. To a certain extent, the researcher can compensate for the lack of experimental control by the use of statistical control. For instance, a researcher might wish to correlate the amount of GUARD FIST I training with live-fire performance on Table VIII using crews from an Army National Guard unit. A problem is that, in an uncontrolled training situation, events other than GUARD FIST I can affect Table VIII performance, for instance, training on M-COFT, VIGS training, or the tank itself. Using techniques such as partial correlation analysis, the effects of these "other" training events can be held constant to isolate the relation between GUARD FIST I and Table VIII. Alternatively, a causal or path analytic model of Table VIII performance can be used to trace relationships among these various determinants of proficiency. This example highlights an important advantage of the large-sample correlational approach: that the training resources are free to vary as they do in the "real world." As a result, the resource-proficiency tradeoffs obtained from these investigations are more generalizable to actual training situations.

This initial description of large-sample correlational designs would make it appear that they offer the best of all worlds: realistic resource-proficiency relationships that can be determined with minimal interference to ongoing training. Furthermore, given an adequate sample size, statistical methods can be used to control for or to simply examine the effects of potentially confounding and/or intervening variables. However, appearances can be deceiving, especially when considering potential problems in actually performing this sort of research. First, the researcher must be able to identify these "other" sources of variation and then devise operations for measuring them. This process requires a fairly detailed knowledge of gunnery training. Second, the amount of training on the resource in question must show some variation. Variability in training resources may not be obtained in units where training schedules are fairly rigid and standardized across

soldiers, crews, platoons, and so on. Clearly, without sufficient variation in training resources, the correlational approach is not a viable alternative. Third, the argument that results from a large-sample correlational design are more generalizable may not be relevant to the researcher's objective. To determine the full range of resource-proficiency effects, for instance, the researcher might be interested in measuring performance under contrived situations that are sometimes better than, and sometimes worse than, the normal state of affairs. Finally, because the relationships are expected to be curvilinear, linear regression techniques may not be appropriate. Analysis is discussed at length in the following chapter.

An example of the correlational approach is Campshure and Drucker's (1990) research on the prediction of live-fire gunnery performance on Table VIII from U-COFT training data. Results from their first set of data provided by two armor battalions (77 crews) indicated that position in the U-COFT training matrix (as defined by Reticle Aim Level) was significantly predictive of first-run Table VIII performance. They found even better prediction of Table VIII total score when Reticle Aim Level was combined with time in crew or with total number of exercises in multiple regression equations. Both time in crew and total number of exercises were weighted negatively in the regression equations, but were not significantly related to total Table VIII score in the bivariate correlations. Thus, these two variables partialed out (or suppressed) that part of the Reticle Aim variable that was unrelated to the total score, thereby improving the predictions. These suppressive relationships suggested that speed of progress through the U-COFT matrix improved prediction over the position per se.

Campshure and Drucker's (1990) results also illustrated the difficulty in interpreting correlational data. In an attempt to replicate their findings, they examined similar data from four additional battalions (136 crews). None of the relationships revealed in the first data set were detected in the second set. The authors speculated that the discrepancy between the findings was due to the amount of on-tank gunnery training that the units were allowed to conduct: The battalions in the first set were under severe tank mileage constraints thereby limiting the amount of on-tank training that they could conduct; in contrast, the battalions in the second set were under no such constraints and therefore participated in virtually unlimited on-tank exercises. Campshure and Drucker argued that the additional on-tank training received by the battalions in the second data set reduced the relationship between U-COFT and Table VIII. Unfortunately, these effects were unanticipated; therefore, measures of on-tank training were not obtained. This failure points out the need to identify the principle determinants of performance prior to the research and to devise operations for quantifying each variable.

Indirect-Control Designs

As an alternative to direct manipulation of training resources as prescribed by experimental designs, the experimenter can exert indirect control over training resources through soldier selection or by the employment of comparison groups. In a sense, indirect control is a compromise between total control of training resources as prescribed by the experimental designs and the lack of control in the pure correlational designs.

Backward transfer. One approach to validating training devices is to select soldiers who differ in the skills purportedly trained on the device. The fact that experts perform better on a device than novices is interpreted as evidence that operational skills have transferred to the training device. Thus, this design is sometimes referred to as "backward transfer" (Caro, 1977). Positive results from backward transfer experiments are used as evidence that a device is training the appropriate skills. This reasoning is analogous to the approach used to concurrently validate a newly developed knowledge test wherein performance on the to-be-validated test is compared between two groups--a group that should have the knowledge in question, and one that should not have the knowledge. Differences between groups provide evidence for the validity of the test.

One advantage to the backward transfer approach is that the experimenter can assess the effects of high levels of training and expertise that would be impossible to duplicate in a short-term experiment. That is, the backward transfer experiment examines the effects of years, even decades, of training; whereas the typical transfer experiment examines the effects of hours, or (at most) weeks, of training. On the other hand, the fact that expertise is not under strict experimental control implies that experts and novices may differ on dimensions other than expertise alone. For instance, those individuals having high levels of military experience are survivors of a complex personnel attrition process that selects for appropriate learning aptitudes and attitudes as well as the target knowledges and skills. Such factors could potentially contaminate the resource-proficiency relationship.

For resource-proficiency problems, the most troublesome deficiency with the backward transfer paradigm is that it does not provide tradeoffs that are useful to trainers. For instance, suppose that master gunners perform better than entry-level gunners on U-COFT. In one sense, this information is irrelevant because the military does not invest hundreds of thousands of dollars to develop high levels of expertise so that gunners can perform well on training devices. On the other hand, the results from a backward transfer experiment may be important in some limited cases, particularly in the early stages of device development. The backward transfer design might provide some basic information on the commonality of skills trained on the device and skills required to operate the actual equipment. In that sense, backward transfer is similar to the simplest experimental designs in that it provides only basic information about learning and transfer.

Graham and Smith (1990) employed a variation of the backward transfer design in their investigation on the acquisition of gunnery skills. They tested individual soldiers on the M1 Institutional Conduct-of-Fire Trainer (I-COFT) using three sets of scored exercises: (a) single stationary targets at long ranges, (b) multiple stationary targets at short ranges, and (c) single moving targets at long ranges. They compared the performance of two groups of soldiers: (a) participants in the Excellence in Armor (EIA) program, who were high ability entry-level soldiers selected to receive additional training after completing standard basic training in armor; and (b) senior non-commissioned officers (NCOs), who served as gunnery instructors at the Armor School. The EIAs were tested twice (before and after the additional training), whereas the NCOs were tested once. Results from the performance tests showed that the EIAs significantly improved from pre- to posttest with no differences between EIA posttest performance and NCO performance in accuracy measures for single and multiple stationary targets

presented at short ranges. The researchers interpreted this finding as indicating that accuracy skills required for these targets develop quickly. For long range moving targets, the NCOs were significantly more accurate than the EIAs, suggesting that these skills develop more slowly. With respect to speed of performance, the results showed differences between EIAs and NCOs for all three types of gunnery engagements. This latter finding suggested that skills related to speed of performance develop more slowly than skills related to accuracy. Although this research did not address the resource-proficiency problem per se, it did provide basic information on the time course of gunnery skill acquisition.

Intact comparison groups. If training resources cannot be experimentally manipulated and are not expected to vary substantially within a unit, an alternative is to make comparisons among units that are expected to vary in training resources. This sort of design is used when certain units are selected to test a prototype training device or method. Their performance is compared to the performance of units that are not selected to receive the device. This usually means that the experiment will typically provide only a two-point comparison of training resources--that is, some fixed amount of training vs. no training on the new device.

A commonly recognized problem in this sort of research is the possible confounding due to preexisting differences between intact groups. The severity of this problem is lessened if this design is replicated across multiple units. Another approach is to pretest both groups to determine whether or not there are in fact preexisting differences. If there are differences, techniques (e.g., analysis of covariance) can be used to reduce statistically the effects of initial differences on final performance.

An example of an intact groups design is provided by Morrison and Walker (1990) who examined the effects of mental practice on gunnery performance. These researchers instructed a platoon of entry-level soldiers to use mental practice techniques as they were learning gunnery on the Institutional Conduct-of-Fire Trainer (I-COFT). Performance of this group was compared to performance of another platoon that did not receive mental practice instruction. The gunnery performance of both groups was measured at the beginning and the end of I-COFT gunnery training. There was a nonsignificant trend for the platoon that received mental practice instruction to perform better than the comparison platoon on the posttest. However, there were even larger differences between groups on the pretest. Analyses of covariance were performed to correct for pretest differences, but they revealed no differences between adjusted posttest means. Thus, the researchers failed to demonstrate a proficiency advantage for mental practice. However, they noted that the initial high proficiency of the mental practice platoon, in contrast to the comparison platoon, may have masked any performance gains due to mental practice. With regard to design issues, this research illustrates how preexisting differences between intact groups can make resource-proficiency relationships ambiguous.

Choosing a Correlational Design

As argued earlier, the choice of experimental designs is primarily driven by the issues that the researcher wishes to address. In contrast, the choice of correlational designs is primarily driven by constraints to the research situation. Furthermore, as argued below, the final choice of design

is likely to be an amalgam of correlational and experimental approaches. Consequently, the design-by-issue matrix in Table 1 would not be appropriate. Instead, we offer a some general guidelines on implementing aspects of correlational designs.

<u>Develop model of resource-proficiency relationships</u>. In complex training situations, particularly those over which the researcher has little or no control, there are multiple determinants of gunnery proficiency that are operative along with the resource(s) in question. It would be useful to develop a cause-and-effect model of those determinants, much like structural models used in path analysis. We maintain that a cause-and-effect model helps the researcher develop a conceptual map of training events, even if path analytic techniques are not appropriate to the quantity and/or quality of the data collected. The value of constructing this model is that it helps identify what variables should be controlled and/or measured in order to reach valid conclusions about the resource-proficiency relationship in question.

Figure 10 provides an example model of gunnery training that is intended to prepare crews for qualification on Combat Table VIII. The model illustrates gunnery training that starts on the U-COFT device and on basic Combat Tables I-IV, which are often conducted as dry-fire exercises with tank-appended laser devices. In this hypothetical situation, the two training methods are not connected with a directional arrow indicating a strict cause-effect sequence, because training on these two methods is completely intermixed. Instead, the two training methods are connected with a double-headed arrow indicating a correlational relationship. In contrast, training on these two methods is strictly sequenced to occur prior to but not during the gunnery density, the period immediately preceding Table VIII. Gunnery density begins with a combined Combat Table V/VI, which provides intermediate-level training on live-fire, machine gun, and main gun engagements. Intermediate gunnery training then proceeds to Combat Table VII, which serves as a warm up for Table VIII and is essentially identical to it. If the researcher were interested in examining the impact of U-COFT on live-fire gunnery proficiency, this model clearly points out that effects of U-COFT on Table VIII (the terminal objective) are mediated through basic Combat Tables I-IV and intermediate Tables V-VII. Given this situation, the researcher should supplement data on Table VIII performance with data from preceding tables.

<u>Control primary training resources</u>. In general, the experimenter should, to the extent possible, control the training resources under investigation. The best option is to randomly assign individual soldiers, crews, or platoons to differing amounts of device training. If random assignment is impossible, the second best option is for the experimenter to indirectly control resources through the use of intact comparison groups. This assumes that the intact groups vary in amount of device training. If not, the only alternative is to use the least desirable approach: to simply measure training resources as described in the pure correlational design. A significant problem for uncontrolled correlation of training resources and subsequent proficiency is that training may be allocated by crew proficiency. That is, the least proficient crew may be given more training. Unless a large amount of training is given, the resulting correlation between amount of training and subsequent performance may be negative.

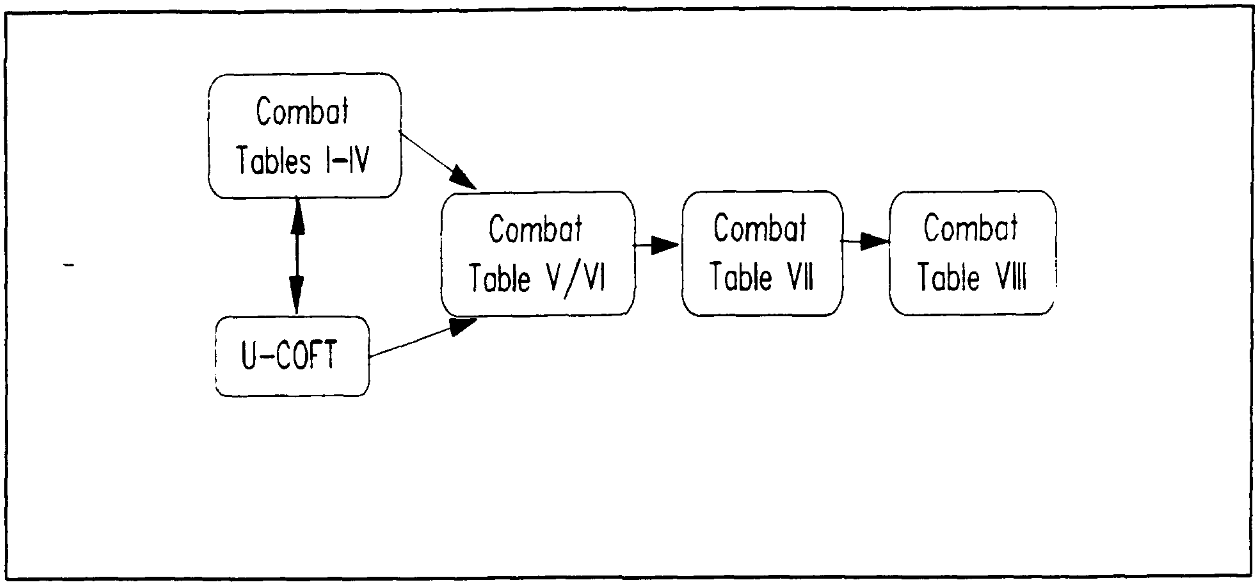

Figure 10. Example cause-and-effect model of gunnery performance determinants.

Measure secondary training resources. To be used to their best advantage, correlational variables should be used in combination with experimental variables. For example, suppose a researcher were interested in the effects of amount of SIMNET training on performance on Tactical Table I. If possible, the researcher should randomly assign platoons to conditions that prescribe differing amounts of SIMNET training. This manipulation would make the effects of SIMNET training independent of other variables; however, the effects of other variables might cloud the relationship between SIMNET and proficiency. For instance, on-tank training on the preceding Tactical Tables (A-H) is designed to prepare platoons for Table I. The researcher might elect to measure the amount of training and the resulting proficiency levels of platoons on each of these preceding tables. These data could be treated as correlational variables along with amount of SIMNET training, which would be under experimental control.

Sample sizes. If a multivariate correlational design is used, the researcher should obtain as large a sample as possible. If relationships are strong and measures highly reliable, the standard wisdom is to obtain no fewer than 10 replications (individuals, crews, platoons) for each statistical parameter to be tested. In tank gunnery research, however, a number of considerations argue for even larger sample sizes. These considerations may be briefly summarized as follows: (a) Within-subject variability makes reliability low (Hoffman, 1989), (b) between-subject variability weakens the practice/proficiency relationship, (c) sampling crews by intact units (e.g., battalions) rather than randomly distorts relationships (Boldovici, 1987; Hoffman, 1989), and (d) mathematical models of skill acquisition and transfer greatly add to the number of parameters to be tested. Given a moderately complex model, the minimum requirement will not be sufficient and many more subjects will be required. Suffice to say that the researcher should obtain as many crews as the research support system will allow.

Recommendations

This chapter has discussed a wide variety of designs that are suitable for issues related to resource-proficiency relationships and training method tradeoffs. The designs were divided into experimental and nonexperimental designs with some general strategies for choosing designs within those two categories. It has also been suggested that other specialized and hybrid designs are possible. Despite the fact that the particulars of research designs for determining tradeoffs can be complex, the basic attributes of the appropriate design can be summarized by a short list of minimum requirements and desirable options. These attributes are summarized in Table 3 and are briefly discussed below.

Table 3

Attributes of Designs for Tradeoff Research

Design Attributes
Minimum Requirements
• variation in resources under investigation (e.g., a training device).
• repetition on criterion training method (e.g., operational equipment).
• other relevant conditions either held constant or measured.
Desirable Options
• experimental control of resources under investigation
• lots of cases (individuals, crews, platoons)

The minimum requirements include those conditions that *must* be fulfilled to measure tradeoffs. An important minimum design requirement is that there be variation in the resources that are under investigation. Variation in a resource (e.g., a training device) may be achieved through experimental manipulation or natural variation in usage. Another requirement is that there be repeated training trials on the criterion training method. The criterion method usually represents the terminal performance on operational equipment, but may (for certain problems) be performance on a training device that is more expensive and/or has higher fidelity than the resource under investigation. The final requirement is that the conditions that potentially impact on criterion performance be held constant or that they be quantified as another resource. If a researcher were interested in the tradeoff between training on a device and the tank itself, for instance, gunnery training by other methods must be considered. The amount of time on the other devices must be held constant for each crew, or, at minimum, the amount of training on the other devices must be recorded.

The desirable options include those conditions that make the determination of valid tradeoffs more likely. The first option is that the resource under investigation be placed under experimental control rather than be allowed to vary naturally. This option ensures that resources vary in a systematic and predictable fashion. The second such option is to obtain measures from as many experimental cases as possible. Depending on the gunnery problem, these cases may be individual soldiers, tank crews, or armor platoons. Large samples make the design more sensitive to actual relationships and tradeoffs among the training resources.

Chapter 3. Data Analysis Methods

In addition to attending to design issues, research on resource-proficiency relationships and tradeoffs requires careful consideration of data analysis. Assessing tradeoffs is more complex than estimating the simple linear differences in performance between two or more groups that have been trained by different devices. Two important factors, elaborated by Roscoe and Williges (1980) and Bickley (1980), create the need for more complicated models. First, comparisons concerning proficiency resulting from one level of resource expenditure may not apply at other levels because practice does not typically result in a linear increase in performance. Rather, performance tends to improve in increasingly smaller increments as illustrated by the familiar negatively accelerated learning curve. Second, tradeoff research concerns more than differences in performance. It also implies an assessment of the relative efficiency, in terms of training time or costs, by which the performance differences are achieved. Thus, training with Method A may result in higher performance than training with Method B, but if training with B is less expensive or less time consuming, conclusions regarding training are not straightforward. Choosing a mathematical model to guide the data analysis becomes something very different from the standard research convention of analyzing data with techniques based on the general linear model, such as the analysis of variance (ANOVA). Moreover, data analysis for tradeoff problems is not simply a matter of fitting learning curves from the data.

The following discussion provides the background to this problem and presents techniques for solving tradeoffs where the transfer relationships between devices are complex. By necessity, the discussion is mathematical. Analyzing tradeoffs cannot be reduced to the kind of qualitative, usually dichotomous, decisions made under typical hypothesis testing research (i.e., whether or not an effect is statistically significant). Rather, analyzing tradeoffs is more like using multiple regression models to make predictions. However, the nonlinear equations used to make tradeoff predictions are much more complex, so much so that obtaining a solution is not guaranteed. It may be noted that the references cited have generally only skirted the issues addressed below, and only one example was found in the literature of actually making tradeoff analyses based on empirical data (Bickley, 1980). The lack of empirically based examples may attest to the difficulty of the process.

Figure 11 depicts an overview of how the topics in this chapter fit the overall process of making tradeoff decisions. The process begins jointly with consideration of data collection methods and selection of a mathematical model for the data analysis. To repeat, the selection of a mathematical model in traditional research is usually automatic--the general *linear* model is the norm. However, to analyze learning and make comparisons between methods of learning, the linear model will not suffice. Relationships between training and performance are more likely to be nonlinear. Furthermore, the issue is not simply to track learning by one training method, but to make comparisons of learning for different training methods. The majority of this chapter concentrates on specifying equations that can be used to make such tradeoff comparisons. There are a variety of options available, none of which are simple. Once a type of equation (or types of equations if more than one is needed) has been selected, the next step is to use the data to estimate parameters for the equation. As indicated above, this will likely involve the use of nonlinear data analysis techniques. The chapter does not describe how to conduct these analyses. Whatever computer package the analyst uses will

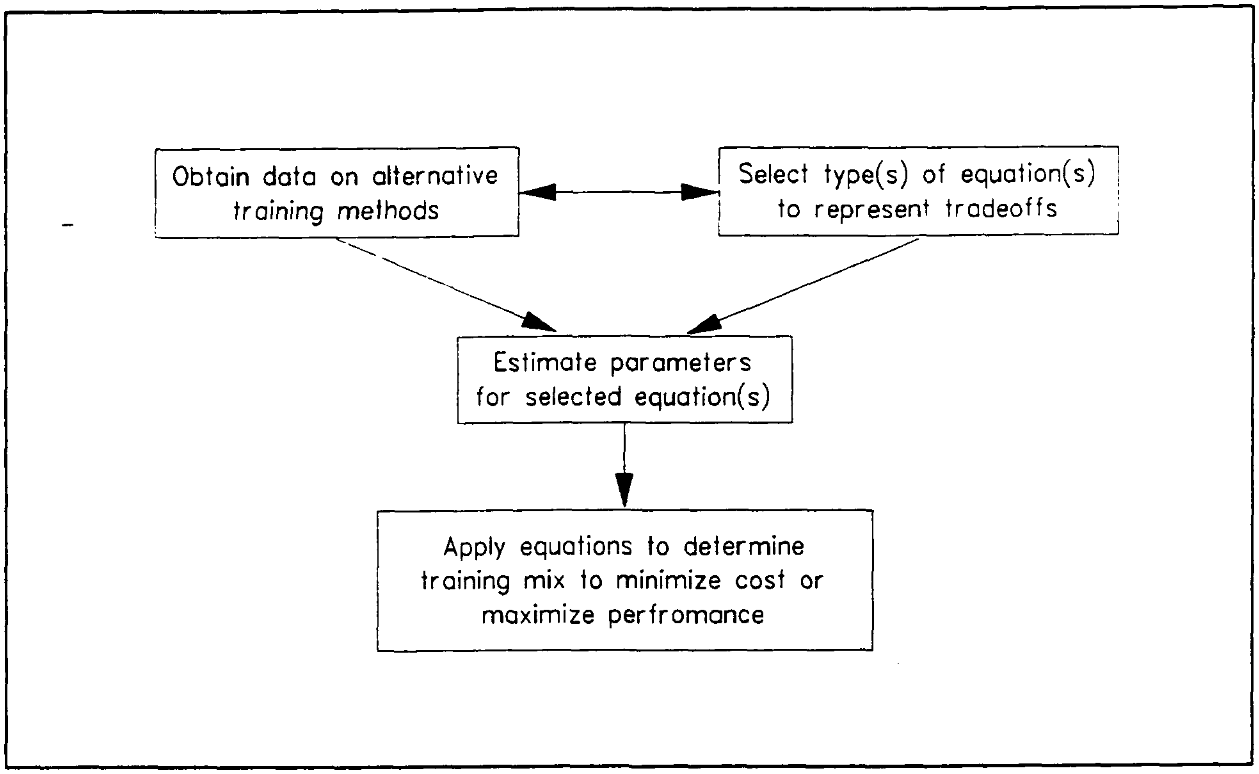

Figure 11. Overview of steps needed to make tradeoff comparisons between alternative training methods.

contain basic information and references. It is important to understand that nonlinear solutions are uncertain and the results may be unstable. Finally, having derived parameters for an equation, the final step is the relatively trivial step of applying the equation. That step may be as simple as plotting the curve and reading from the graph the amount of device training at which lowest cost expenditures are achieved.

The reader may approach this chapter at several levels. The chapter spells out details of very complex formulas. For the most depth, the reader may want to analyze these formulas. Without much loss in understanding, the reader may skim over the formulas and digest from the narrative what the formulas are trying to accomplish. At the simplest level, the reader may want to skim the entire chapter, recognizing that empirically deriving equations to determine resource tradeoffs between alternative training device is highly complex.

Mathematical Models of Resource-Proficienc Tradeoffs

As stated in the Chapter 1, resource-proficiency tradeoffs are basically comparisons of resource-proficiency relationships (i.e., learning curves). There exist several variations on the same theme of how to use learning curve formulations to solve problems related to mixing and substituting different training methods (e.g., device vs. the actual equipment) to achieve the greatest proficiency with the fewest resources.

Direct Estimation of Iso-Performance Curves

The solution used by Povenmire and Roscoe (1973) and Bickley (1980) to analyze training method tradeoffs was to directly estimate savings in training on the actual equipment, the aircraft, as a function of pretraining on a flight simulator. Students began training on a flight simulator and finished in the aircraft. The researchers manipulated amount of training time on the simulator and then recorded the amount of training time in the actual aircraft needed to meet a predetermined standard of performance. For a number of different aircraft tasks, Bickley was able to produce exponential "iso-performance" curves that showed amount of aircraft training needed to reach a predetermined performance standard as a function of amount of simulator training. This function describes all possible mixes of device and aircraft training that would produce the performance standard. The general shape of an iso-performance curve was presented as Figure 7 in Chapter 2. Bickley also showed that by appropriately combining iso-performance functions with cost functions (i.e., costs per training trial), it is possible to identify an optimum mix of simulator practice and aircraft practice that minimizes total costs. The result is a total cost curve, as illustrated in Figure 12, that clearly shows the amount of device training that leads to the lowest total cost of training.

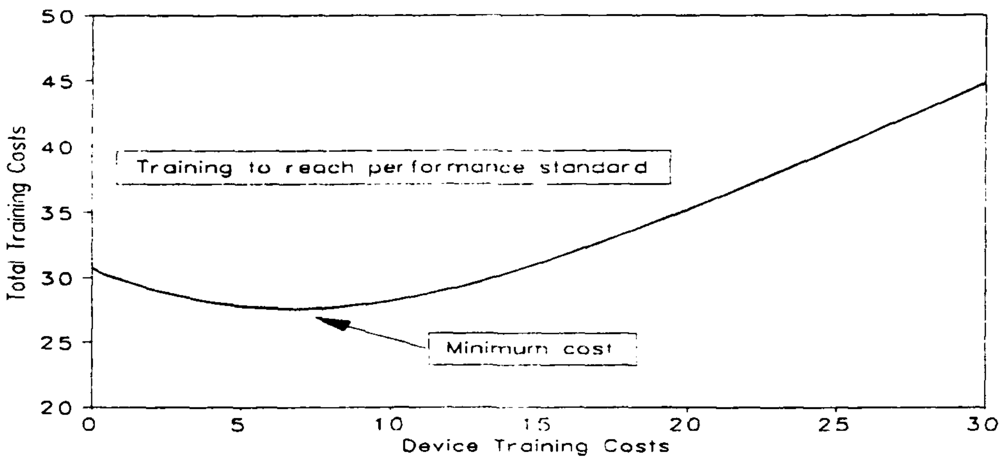

Figure 12. Sample total cost curve.

Carter and Trollip (1980) presented essentially the same idea; however they emphasized that the terminal performance level is not always fixed. For each alternative performance standard, there is a separate iso-performance curve. These curves are related such that any one curve may be considered as part of a family of iso-performance curves, each curve relating actual training to device training for a different standard of terminal performance. Based on this family of curves, Carter and Trollip developed an analysis method for determining the mix of device and operational equipment training which maximizes terminal performance for a given training budget. Thus, the iso-performance curve concept is pivotal in analyzing training tradeoffs, particularly if any one curve for a pair of training methods (e.g., a device

and the operational equipment) is viewed as representative of a family of curves for that pair of methods.

Mathematical Construction of Iso-Performance Curves

Cronholm (1985) conceptualized the training analysis problem more generally. His concept was to develop a mathematical model that integrates learning curve formulas for different training approaches. The model can then be used to solve training mix problems to either minimize costs or maximize performance. His discussion focused on training two different tasks for which practice of Task 1 enhances performance on Task 2.[4]

Cronholm's (1985) approach requires basic learning curves (i.e., resource-proficiency relationships) for each task showing task performance as a function of training time or trials on that task. The approach also requires a function that relates Task 1 performance to Task 2 performance. This latter function is then used to mathematically transform the basic Task 1 learning curve into a learning curve that indicates Task 2 performance as a function of Task 1 training. Thus, two curves exist for Task 2 performance, one as a function of Task 1 training and another as a function of Task 2 training. For any given number of Task 1 training trials, the expected level of Task 2 performance can be determined from the first curve. Then, assuming the Task 2 training begins at the level of performance at which Task 1 training ends, the second curve shows the amount of additional Task 2 training for meeting various levels performance. Thus, the two curves can be used to construct iso-performance curves for any given standard of performance. (The mathematical transformation of the learning curves into a iso-performance curve is presented later in this chapter.)

Cronholm (1985) also showed that by using a simple ratio relating the cost of Task 1 training to cost of Task 2 training, iso-performance curves can be constructed in terms of equivalent training costs rather than training time or trials. That is, a converted iso-performance curve shows how a given level of performance can be achieved by various allocations of training resources. A total cost curve can then be constructed and used to solve either cost minimization or performance maximization problems.

Finally, Cronholm (1985) showed that the amount of Task 1 training that minimizes training cost is the same regardless of the performance standard required for Task 2. This is an important idea for three reasons. First, optimization results from Roscoe's (1971) research design, which requires a preset training standard, can be generalized to other terminal performance standards. That is, regardless of the terminal standard for operational equipment performance, there is a particular amount of device pretraining that is optimal and this amount can be found from a study that investigates only one level of standard performance. Second, because the amount of device pretraining that yields minimum cost for any given performance standard is

---

[4]In general the primary concern in resource-proficiency tradeoff research is the tradeoff between a training device and the operational equipment. However, the tradeoff concepts being presented in this chapter are applicable to any set of training methods that are intended to have a positive transfer effect on each other.

constant for all performance standards, the same amount of device training will maximize performance for any fixed training budget. Finally, it follows that a training program does not have to have a hard-and-fast standard of performance in order to optimize device use.

Analysis of the Slopes of Training Cost Functions

Sticha, Schlager, Buede, Epstein, and Blacksten (1990) also presented a mathematical model for determining optimal training mixes. They were interested in the cost savings that might accrue if underlying skills were trained prior to training complete tasks. By substituting "device" for "skill" and "task" for "operational equipment," the mathematical concepts presented by Sticha et al. also apply to problems of training on devices prior to training on the operational equipment.

Similar to Cronholm (1985), Sticha et al. (1990) began with a learning curve function that relates amount of skill training to skill proficiency (analogous to Cronholm's Task 1), and combined it with a transfer function that gives task proficiency (analogous to Cronholm's Task 2) as a function of skill proficiency. The result was a function that gives task proficiency as a function of skill training. It is at this point that Sticha et al.'s approach diverged from Cronholm's. Sticha et al. converted skill training trials to skill training costs in the basic learning curve so that skill proficiency is expressed as a function of training costs rather than training trials. To simplify cost estimations, skill training costs are not expressed in dollars but in terms of the amount of resources needed for one unit of task training. Thus, one cost unit equals the resources needed for training one trial on the operational equipment. Using this cost metric, the learning curve relating task training trials to task proficiency becomes a cost function as well; therefore, it is possible to compare task proficiency expected from resources spent on task training to task proficiency expected from the equivalent resources spent on skill training.

Sticha et al. (1990) proposed analyzing the relative costs of task and skill training by comparing the slopes of the task training and the cost-converted, skill training curves. Because the curves are expected to be negatively accelerating, the slopes of both curves change (i.e., diminish) with increasing training and proficiency. However, the slopes of the two curves are not expected to change at the same rate. Sticha et al. (1990) argued that at any given level of proficiency, the most cost efficient training method is the one providing the greatest increment in task proficiency--that is, the one with the steepest learning curve slope. The slope for skill training is expected to start out higher than the slope for task training, but is also expected to diminish more rapidly than the slope for task training. Thus, skill training (or device training in our case) may increase proficiency at a faster rate per unit of training cost at the beginning levels of proficiency, but task training (or operational equipment training) may increase proficiency at a faster rate per training cost at the later levels of proficiency. They suggested that the optimum mix of training is defined by the level of performance at which the slopes of the learning curves are equal. Thus, Sticha et al. compared cost function slopes instead of constructing iso-performance and total cost curves to identify optimum training mixes. Their expectation was that the point at which their learning curve slopes are equal will identify the same amount of device training identified by the minimum point on Cronholm's (1985) total cost function.

Sticha et al.'s preference for analyzing slopes was based on a desire to more conveniently handle tradeoff problems involving more than two training alternatives (P. J. Sticha, personal communication, March 1991). That is, learning curves for several training methods could be adjusted to a common cost metric and the slopes of these curves compared to determine which training method provided the greatest gain at various levels of proficiency.

Figure 13 illustrates the concept of comparing learning curve slopes to determine optimum training method. Three cost-based learning curves are presented, one each for Methods 1, 2, and 3. The curve for Method 1 has the steepest initial learning gradient, and therefore provides the most cost efficient training up to the proficiency level indicated by point A. At point A, the slope of the Method 2 curve equals the slope of the Method 1 curve; beyond point A the slope to Method 2 is greater than Method 1. Between points A and B, the slope of the learning curve for Method 2 is also greater than the slope for Method 3. Therefore, between proficiency levels A and B, Method 2 provides the most cost efficient training. Beyond point B, the learning rate for Method 3 is greater that either Method 1 or 2. Therefore, if training were to begin for novice students, Method 1 should be used until the students reach point A. Assuming that Method 3 is used as the cost metric, the amount of training on Method 1 needed to reach point A is equal in dollar amount to approximately 2 units of Method 3 training. At this point training would shift to Method 2, as signified by the horizontal arrow, and students would now be progressing up the Method 2 learning curve. To improve from proficiency level A to level B would require Method 2 training equivalent in cost to approximately 7 units of Method 3 training (12 - 5 on the X-axis). At level B, training shifts to Method 3, and students would be progressing up the Method 3 curve. Notice that the cost savings of combining training methods are indicated in the figure as the total lengths of the two horizontal arrows. That is, by using the recommended sequencing Method 1 saves 3 units of cost in reaching 50% proficiency and Method B saves 8 units of cost in reaching 70% proficiency, for a total of 11 cost units saved over using Method 3 alone.

The figure also illustrates two other important principles. First, it shows that device optimization is independent of any performance standard. Regardless of the final proficiency target, costs are minimized by using Method 1 until proficiency reaches approximately 50%, followed by Method 2 up to a proficiency level of 70% and then completing training on Method 3. The second principle is related. Answering how much training should be spent on one training method relative to another depends on the student as well as the learning curve characteristics for the alternative training methods. Using Figure 13 as an example, novice students should train on Method 1 for about 2 cost units worth of time or trials. On the other hand, more advanced students, who are more than 50% proficient, should not use Method 1 at all. Thus, questions about alternative resource investments must always consider the background of population to be trained.

A Critical Assumption in the Methods for Estimating Tradeoff Relationships

The approaches of Cronholm (1985) and Sticha et al. (1990) require three functions as input to their mathematical models: (a) a learning curve for device pretraining, (b) a learning curve for criterion training (e.g., operational equipment training), and (c) a function relating device performance to operational equipment performance. Consider the groups-by-trials design presented in Chapter 2 with four groups, each with four blocks

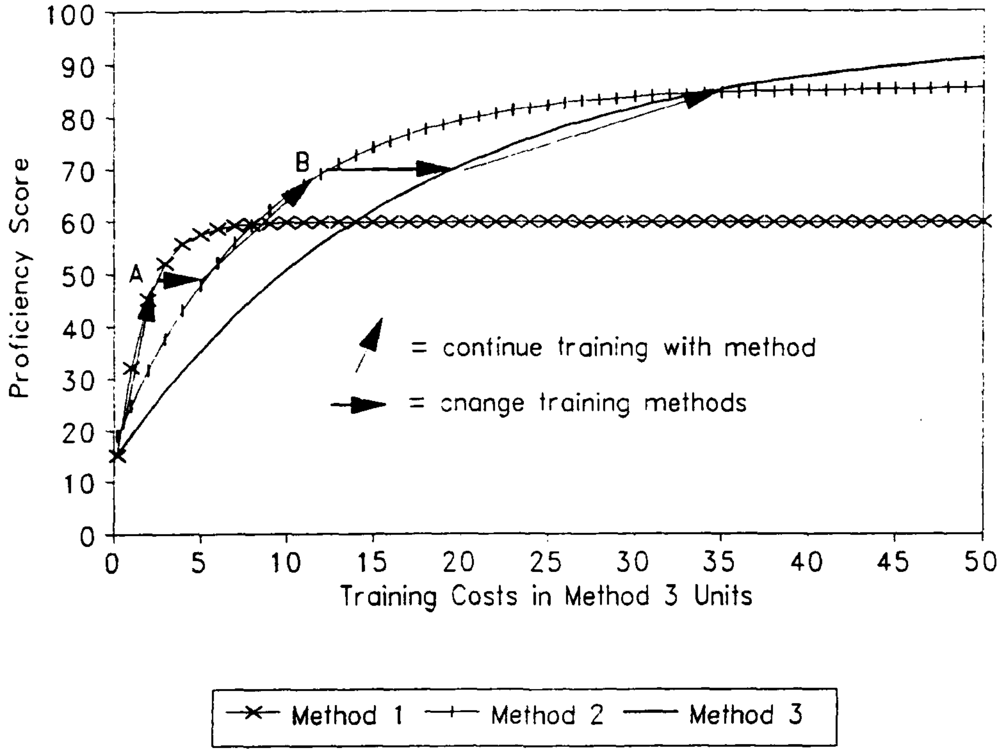

Figure 13. Training methods selected to maximize performance gain: Train with Method 1 to Point A, then with Method 2 to Point B, and finally with Method 3.

of equipment training, but with four different amounts of device pretraining. This design provides a data set from which all three curves can be produced. In addition, the design offers a shortcut. Cronholm and Sticha et al. mathematically constructed a learning curve with operational performance as a function of device training. Using the groups-by-trials design, this curve can be estimated directly using the variation in trials across groups and operational equipment training score from the first trial.

Herein lies the critical assumption of the Cronholm (1985) and Sticha et al. (1990) models. The group-by-trials design was suggested because of the weakness of focusing on first trial performance on the operational equipment following device training. Their manipulations of learning functions assume that learning on the operational equipment may be depicted by the same learning curve whether or not there is prior training on a device. As illustrated above, learning is depicted as movement along a learning curve. Switching training methods during the course of learning is modeled as switching from one learning curve to the other, *beginning on the operational equipment curve at the same performance level at which device pretraining was terminated.* There is no provision for "interfacing" (see Chapter 2), which would be represented as performance on the operational equipment beginning below that expected from final device performance. Neither is there any provision for potential synergistic interaction effects where training

improvement based on combinations of training methods are greater than any one method alone. These two effects would have to be modeled as a changes in the shape of the operational equipment training curve. Nevertheless, the models of Cronholm (1985), and Sticha et al. (1990) are instructive in terms of telling us the kind of information and mathematical manipulations needed to reach tradeoff decisions to minimize costs or maximize performance.

Notice also that Bickley's approach of directly estimating iso-performance curves skirts the above issue. Any interfacing or synergistic effects that might be occurring in the instructional process are incorporated in the estimates of the number of equipment trials that are needed to reach standard after device training has been completed.

Moderated Mathematical Models of Learning with Multiple Training Methods

Clearly, there are a number of difficult issues that have not been addressed by Bickley (1980), Carter and Trollip (1980), Cronholm (1985), or Sticha et al. (1990). These issues concern the uncertainty of how pretraining by one method affects both initial performance and subsequent learning on other training methods. In this section we will explore some additional modeling issues pertinent to training tradeoff analysis which include the possibility that device pretraining changes the shape of the operational equipment learning curve. At the outset, we should indicate that basically we are between a rock and a hard place. Mathematical solutions quickly become so complex that without a large amount of empirical data the ensuing analyses will be unstable at best. Thus, we are led into a position of needing complex mathematical models (the rock) and needing a large data base to solve them (the hard place). Undoubtedly this will lead to compromise solutions.

Alternative Analytic Approaches

Consider again the data that would be provided by the groups-by-trials design. There are several alternative levels of sophistication with which one could approach the analysis of such data given expectations that curve shapes may be different for the different treatment groups. At the simplest level, an ANOVA could be conducted, using procedures for repeated measures with a single grouping factor. The four levels of device training define the grouping factor and equipment training performance is the repeated trials effect. If pretraining affects operational equipment performance in any nonlinear manner (e.g., performance improving faster during early equipment trials), then one would expect a significant main effect for group as well a significant trials effect and a group-by-trials interaction. We may even expect significant second- and third-order polynomial contrasts for the trials and group-by-trials effects due to nonlinear trends in the data. Obviously, positive results would be indicative of some type of pretraining effects that could be plotted and graphically interpreted, but this analysis would not satisfy mathematical requirements for constructing resource-proficiency tradeoff functions. Although the results of the analysis may indicate that pretraining effects are present, they do not yield learning curve formulas that can be transformed into the iso-performance curves needed for tradeoff analyses.

At a second level of sophistication, one could fit a curvilinear function (alternative functions are described in the following section) to the operational equipment performance data for each of the treatment groups and

then examine differences in the curves by comparing equation parameters. The operational equipment learning curves could be plotted for each group and subjective differences in curve shapes could be narratively described. Furthermore, the results from nonlinear methods needed to fit the learning curves are inherently unstable. Thus, true differences between curves may be hard to detect. More importantly, there is no compelling way to mathematically combine the different curves into a simple function that can be used to analyze resources-proficiency tradeoffs.

The third solution is to bypass estimation of device and equipment learning curves and directly estimate iso-performance curves. To obtain these curves, one would convert the groups-by-trials data to Roscoe-type data by selecting a level of operational equipment proficiency that is met by all, or nearly all subjects and counting operational equipment training repetitions needed to meet that level of proficiency. An iso-performance curve could then be fit with nonlinear methods relating device training and additional equipment training needed to met the selected proficiency level. (See Figure 7 in Chapter 2.) This is a solution for estimating resource-proficiency tradeoff that avoids the complex formulas that will be presented below. Its disadvantage is that no information is gained about the learning curves per se. Learning curve information is important for making adjustments in tradeoff recommendations to match different trainee populations.

The final level of sophistication is to use moderated mathematical models in which parameters of equipment learning curves are described functions of the amount of device pretraining. This approach will be described below using sample mathematical functions for illustration. First, formulas for four alternative learning curve formulas will presented to describe basic resource-proficiency relationships. Then, these alternative formulas are used to discuss different ways that learning curves could be moderated by pretraining on a different method. These moderated curves, which combine information about two or more training methods, are the basis for resource-proficiency tradeoff analyses.

<u>Alternative Learning Curve Functions</u>

It is well accepted that proficiency increases more rapidly during early practice than during later practice as described by the negatively accelerated learning curve. The curve may either increase or decrease depending on the proficiency measure; for instance, time and error measures have negatively accelerated decreasing curves whereas proportion correct measures have negatively accelerated increasing curves. Several different mathematical functions have been proposed to represent such curves, including power functions, exponential functions, hyperbolic functions, and logistic functions. Lane (1987) also presented variations for each function, including increasing and decreasing forms, with and without selected curve parameters. No one particular curve appears to have universal acceptance, and in practice the differences are slight. For example, after examining over 40 different data sets, Mazur and Hastie (1978) favored the hyperbolic function over the exponential, although the differences in fit between the two forms were negligible in many cases. Much more obvious from Mazur and Hastie is that the data do not fit a linear pattern. Our immediate concern deals less with the particular form that is best than with the use of nonlinear functions for analyzing training data in order to develop projections concerning training tradeoffs.

Instead of working in the abstract as Cronholm (1985) did, we will examine four specific functions that may be used to represent the typical learning curves, focusing on the increasing forms for each. The formulas expressing these functions are presented below:

$$\textit{Power:}\quad Y = A - G(N+1+B_e E)^{-R} \tag{1}$$

$$\textit{Exponential:}\quad Y = A - Ge^{-R(N-1+B_e E)} \tag{2}$$

$$\textit{Log/Log:}\quad Log(Y) = K + RLog(N) \tag{3}$$

$$\textit{Linear:}\quad Y = K + B_1 N + B_2 N^{.5} \tag{4}$$

Variables to be measured:

- Y = Performance (e.g., Tank Table VIII score)
- N = Units of training (e.g., trials, time, or sessions)
- E = Previous experience (e.g., previous Tank Table VIII qualifications, time in position)

Parameters to be estimated:

- A = Asymptote (maximum performance)
- G = Maximum gain in performance (A - G = initial performance after one unit of training and no previous experience)
- R = Learning rate (how steeply the curve rises)
- B_e = Previous experience weight (correction factor that converts previous experience, however it is expressed, into equivalent training trial units)
- K = constant that carries initial performance information
- B_1 & B_2 = regression weights that carry rate and asymptote information

The first two functions, power and exponential, must be analyzed by nonlinear regression methods. Both of these functions have four parameters to be estimated: asymptote (A), a weight for previous experience (B_e), learning rate (R), and initial starting point which is represented by a maximum gain parameter (G). Asymptote and experience are left out of some variations of these formulas. We prefer to include them for the following reasons. First, the asymptote parameter is needed for increasing performance curves because there are ceilings on gunnery performance scores. For example, proportion correct metrics are limited to 100%, and artificial scoring limits are placed on other aspects of performance, such as on Table VIII (Hoffman, 1989). Second, the experience variable and its parameter are included because research in the Army context is frequently conducted with experienced soldiers. Because training experience during research (i.e. "experimental trials") does not coincide with total task experience, a previous experience index is needed to correct for the difference.

The log/log and linear functions are presented because they can be estimated using ordinary multiple regression techniques. The log/log function

is a simplified power function for which there are no asymptote or previous experience parameters. Initial performance information is carried in the intercept constant (K).[5] The linear equation is a quadratic form that creates a parabola. The square root of trials is added instead of trials squared to turn the parabola on its side. This allows approximation of an asymptote without forcing the curve to turn down at higher trials. In this function, the intercept constant (K) carries starting point information, and the two regression weights together contain learning rate.

These formulas provide four common mathematical forms for describing a learning function. Merely capturing learning curves, however, is only an initial step in determining resource-proficiency tradeoffs. We are also interested in how learning by one training method affects learning by another. That is, after various amounts of pretraining on a device, what does the learning curve look like for training on the actual equipment? We will describe below the various effects pretraining could have.

Pretraining as a Moderator of Learning Curves

Training on a device prior to training on the operational equipment can moderate the shape of the equipment learning curve in three different ways: (a) change the initial performance starting point, (b) change the rate of performance improvement, and (c) change the performance asymptote. Thus, device pretraining can potentially moderate any of these three curve parameters. Conceptually, the moderator effect of pretraining on the learning curve parameters is analogous to linear interactions. That is, in a linear model, an interaction of two variables (X_1 and X_2) is expressed as:

$$Y = B_1X_1 + B_2X_2 + B_3X_1X_2 + K \tag{5}$$

which can be rearranged algebraically to:

$$Y = (B_1 + B_3X_2)X_1 + (B_2X_2 + K) \tag{6}$$

In Equation 6, it is apparent that the variable X_2 moderates the relationship between X_1 and Y in two ways. Both the slope of the line ($B_1 + B_3X_2$) and the intercept ($B_2X_2 + K$) are functions of X_2. That is, the moderator effects of X_2 on the relationship between X_1 and Y are expressed in the equation parameters. Similarly, moderator effects of pretraining by one method on the learning curve for a subsequent method can be expressed in terms of changes in the learning curve parameters. That is, learning curve parameters for operational training may be devised that contain indices of the effects of device pretraining.[6] Techniques for their mathematical modeling are discussed below.

[5]In the Log/Log and Linear formulas, the K parameter is not actually the starting point, but it adjusts the curve horizontally so that it runs through initial performance when the amount of training equals one unit.

[6]It may be noted that these moderator effects are themselves linear. While the learning curves are non-linear, we will retain some semblance of parsimony by proposing only linear moderator effects.

Change in curve starting point. One possible effect of device training may be a change in initial performance, as expressed by the maximum gain parameter (G) for power and exponential curves. Figure 14 shows device training raising the initial starting point on the operational equipment by reducing the maximum gain of the operational equipment curve. For the log/log and linear model, moderating K will change the starting point. The equations below show how initial starting point effects can be captured in the sample learning curves, where the subscripts o and d designate operational equipment and device training, respectively.

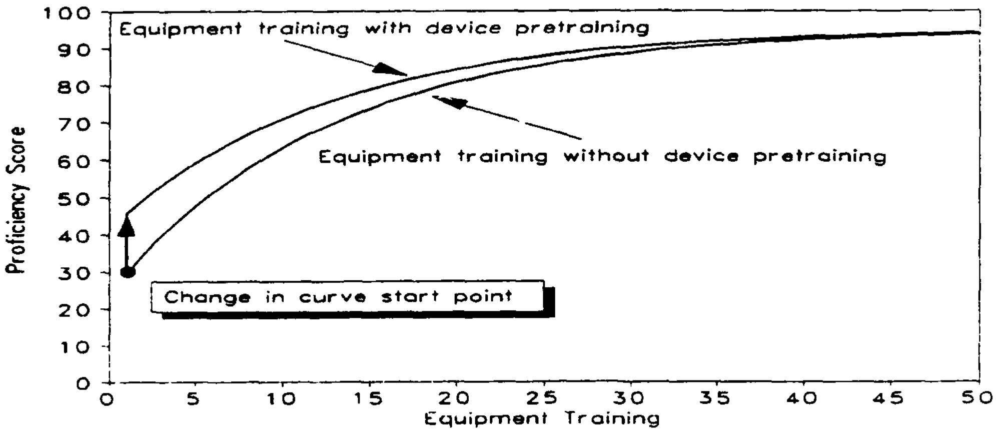

Figure 14. Effects of pretraining on initial starting point of operational equipment learning curve.

$$\textit{Power}:\quad Y = A-(G+B_gN_d)\,(N_o+1+B_eE)^{-R} \tag{7}$$

$$\textit{Exponential}:\quad Y = A-(G+B_gN_d)\,e^{-R(N_o-1+B_eE)} \tag{8}$$

$$\textit{Log/Log}:\quad Log(Y) = K+B_kN_d+RLog(N_o) \tag{9}$$

$$\textit{Linear}:\quad Y = K+B_kN_d+B_1N_o+B_2N_o^{.5} \tag{10}$$

Thus, in the power and exponential formulas, the variable N_d, weighted by the parameter B_g, is added to G. The result is that maximum gain is now expressed by the term:

$$(G+B_gN_d)$$

That is, G in the operational equipment learning curve is moderated, i.e., changed, by amount of device training, N_d, with the strength of the change indicated by the size of the parameter B_g. Amount of device training would be a variable in the data set along with amount of operational training and proficiency score; its parameter B_g would be estimated in the analysis. In the log/log and linear models, change in curve starting point is expressed as a change in intercept by adding B_kN_d.

Change in learning rate. Another hypothesis is that device pretraining may provide a synergistic boost to the effectiveness of operational training. This may be expressed as device training affecting the rate (R) of learning during equipment training. Figure 15 illustrates how a change in rate could change the equipment learning curve. The equations below show how this effect would be captured in the sample learning curve equations. Log/log and linear models are expanded to show that they consist of product terms (like typical interaction terms) so they may still be solved by linear regression.

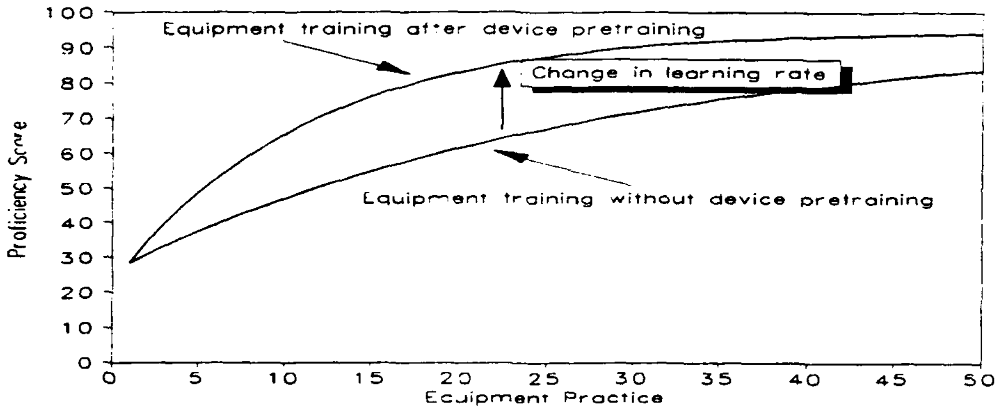

Figure 15. Effects of pretraining on learning rate of equipment learning curve.

$$\textit{Power:}\quad Y = A-G(N_o+1+B_eE)^{-(R+B_rN_d)} \tag{12}$$

$$\textit{Exponential}\quad Y = A-Ge^{-(R+B_rN_d)(N_o-1+B_eE)} \tag{13}$$

$$\textit{Log/Log:}\quad Log(Y) = K+(R+B_rN_d)Log(N_o) \tag{14}$$

$$Y = K+RLog(N_o)+B_rN_dLog(N_o) \tag{15}$$

$$\textit{Linear:}\quad Y = K+(B_1+B_{d1}N_d)N_o+(B_2+B_{d2}N_d)N_o^{.5} \tag{16}$$

$$Y = K+B_1N_o+B_{d1}N_dN_o+B_2N_o^{.5}+B_{d2}N_dN_o^{.5} \tag{17}$$

Change in asymptote. The third potential moderating effect of device training is on performance asymptote. Device pretraining would not normally be expected to change the asymptote of proficiency on the actual equipment, unless (a) the measurement of equipment proficiency excludes some portion of the task, and (b) the device provides practice on that portion of the task. In that regard, U-COFT and GUARD FIST I are able to train greater mixes of targets than those that are included on Tank Table VIII. In order to capture these training effects of U-COFT or GUARD FIST I, proficiency would have to be measured on an expanded performance test. Hoffman, Fotouhi, Meade, and Blacksten (1990) have recommended using a combination live-fire and dry-fire exercise that includes larger numbers of mixed targets as a more complete measure of gunnery proficiency than Table VIII alone. Device pretraining may

moderate the performance asymptote on this expanded criterion such that asymptotic performance is greater with device training than without it. Figure 16 illustrates this effect. The equations below show how these effects could be modeled in the power and exponential forms of a learning curve. With no asymptote parameter, such effects would not be apparent in log/log or linear regression forms.

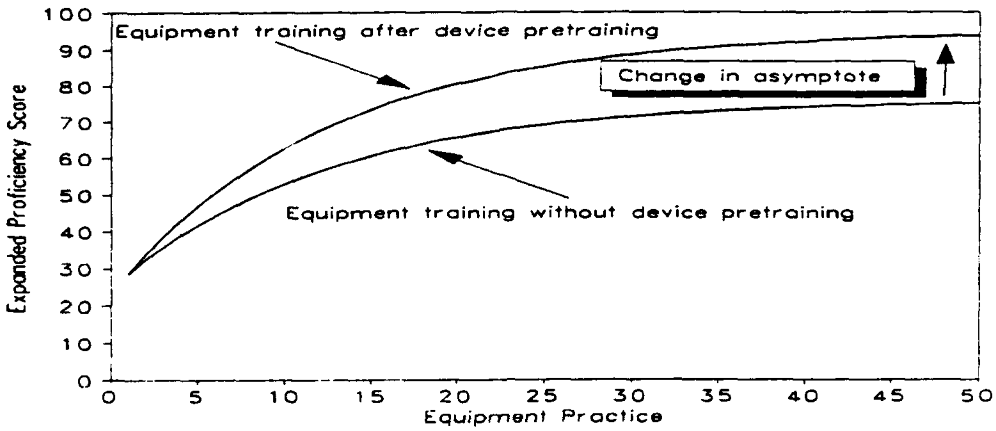

Figure 16. Effects of device pretraining on learning asymptote.

$$\textit{Power:}\quad Y = (A+B_aN_d)-G(N_o+1+B_eE)^{-R} \tag{18}$$

$$\textit{Exponential:}\quad Y = (A+B_aN_d)-Ge^{-R(N-1+B_eE)} \tag{19}$$

Note that in previous and subsequent models, operational training is synonymous with operational performance. That is, operational training consists of repeated performance on the criterion task. In Equations 18 and 19, it is assumed that conditions are added to criterion performance so that it is more inclusive than training performance. Because of this, the groups-by-trials type data will not suffice. Instead, data must come from mixed levels of training on the two training methods followed by testing on the an expanded criterion measure as described in Chapter 2, Extensions Beyond Two Methods.

Estimating learning curves. Device transfer may be captured as a change in the shape of the equipment learning curve is three ways: (a) initial performance may change, (b) learning rate may change, and (c) potential proficiency, i.e., asymptote, may change. With sufficient data, the moderated models described above could be estimated using the groups-by-trials design. With this approach, differences between equipment learning curves can be described in a manner useful for tradeoff analyses.

However, the supporting nonlinear analyses must still be classified as an art (Wilkinson, 1988). The equations above have presented the effects one at a time. Pretraining could affect all three of the parameters. The mathematical formulas needed to model multiple moderator effects are straightforward combinations of the above formulas, however, they are not

simple. If each of the parameters (i.e. asymptote, maximum gain, and learning rate) were moderated, the resulting equation would contain seven parameters for the power and exponential formulas: the basic four parameters plus one each to moderate asymptote, maximum gain, and learning rate. With that number of parameters, data requirements for obtaining stable estimates would be substantial and the possibility of reaching a stable nonlinear solution uncertain. The analysis may have to entail trial-and-error attempts to find reasonable fits to the data with as few parameters as possible. For example, we could assume that device pretraining would have some detectable influence on starting performance on operational equipment. The initial analyses could begin with an attempt to fit a function with moderator effects on G only. Using the log/log model, only three parameters would have to be estimated (K, B_k, and R in Equation 9); power and exponential models would require estimating five parameters (A, G, B_g, B_e, and R in Equations 7 and 8). Examining moderator effects on G and R would require fitting six parameters for the power and exponential functions (A, G, B_g, B_e, R, and B_r in Equations 14 and 15). Alternatively, the simpler log/log model might be used to capture starting point and rate changes with only four parameters and the solution could be derived from regular multiple regression. It is apparent that the data analysis designs share the same constraints that data collection designs have: the more comprehensive the design, the more difficult they become to execute. The simple solutions are a compromise. We need to keep in mind that the goal is to achieve a reasonable representation of patterns of learning in mathematical form so that tradeoff computations can be made. Results may end up being classified as "the best we can do," rather than "fact." However, solving resource-proficiency tradeoff questions requires such functions, and the "best we can do" by following this approach is better than one-shot linear comparisons among training groups.

An Aside: Device Training as Previous Experience

Having described learning curve formulas, it is now possible to revisit Cronholm's (1985) analysis method for combining learning curves to construct iso-performance curves. Underlying his algebra is the assumption that device training does not change the basic shape of the curve, but only where on the curve learning begins. For example, in Figure 13 pretraining by Methods 1 and 2 changed where learning began on the Method 3 learning curve. In the learning curve formulas, this affect is captured as a modification to previous experience. Specifically, E must be rewritten as a function of the amount of device training. The simplest approach would be to include device training as a weighted additive effect to operational equipment trials. The weight would then be interpreted as the ratio of learning value of device training compared to equipment training. However because this ratio is constant across performance levels, the resulting iso-performance functions would be linear rather than curved. This is unacceptable for two reasons. First, it is unexpected from the literature. Second, it makes the training mix question inconsequential: guidance would be to always use the device with the greater weight. Thus, the linear moderating effect is clearly rejected in this case. The alternative is to look at the moderating function as analogous to the base function. That is, for the power function, device trials should affect total trials as a power function, resulting in a "nesting" of a power function within a power function. Similarly, the exponential function would contain an exponential function moderating the training trials parameter. These nested moderating effects would result in characteristic curvilinear iso-performance curves. This is the model assumed by Cronholm (1985) and Sticha et al. (1990)

to describe how pretraining affects subsequent training. Note, however, that the model assumes that the proficiency gained in pretraining is immediately transferred to equipment performance without interfacing or synergistic effects.

The formulas which represent this type of moderating effect are developed in three steps which require some explanation. The goal is to add a function to the trials parameter that gives the number of operational equipment training trials needed to reach the same level of proficiency, Y_d, achieved after N_d trials on the device. Thus, the first step is to translate the basic learning curve function into an equation that gives N as a function of proficiency, Y. For the power function, this is

$$N = \left(\frac{G}{A-Y}\right)^{\frac{1}{R}} - 1. \tag{20}$$

The second step is to add this function to the trials parameter in place of *E*. For the power formula, the result is

$$Y_o = A_o - G_o\left(N_o + \left(\frac{G_o}{A_o - Y_d}\right)^{\frac{1}{R_o}} - 1\right)^{-R_o}, \tag{21}$$

where the subscripts *o* and *d* represent operational equipment and device training, respectively. Next, the device learning curve is substituted for Y_d,

$$Y = A_o - G_o\left(N_o + \left(\frac{G_o}{A_o - A_d + G_d(N_d + B_e E + 1)^{-R_d}}\right)^{\frac{1}{R_o}} - 1\right)^{-R_o}. \tag{22}$$

The analogous formula for the exponential function is

$$Y = A_o - G_o e^{-R_o\left(N_o - 1 + \frac{1 + \ln(E_o) - \ln\left(A_o - A_d - G_d e^{-R_d(N_d - 1 + B_e E)}\right)}{R_o}\right)} \tag{23}$$

These formulas describe performance as a function of learning on the device followed by learning on the operational equipment such that equipment training begins at the performance level at which device training ends. No forms are created for the linear and log/log models because they have no experience parameter. Because of the repetition of parameters in these formulas (e.g., A_o appears twice in each formula), direct estimation of parameters may not be feasible. However, the solution can be accomplished in pieces by estimating parameters for operational training and device training separately. Using data from the groups-by-trials or Roscoe design, performance on the first equipment training trial is the dependent variable for the device curve. The equipment curve is obtained from multiple measures of the group that receives no device pretraining. The two curves are then "assembled" as indicated above.

Using Moderated Models to Solve Optimization Problems

Assuming that moderated models of training can be obtained, it is relatively straightforward to convert the equations to forms useful for tradeoff analysis. The equations can be mathematically converted to iso-performance curves by solving for N_0, the number of practice trials on the operational equipment. For the power function, with the start point and rate parameters moderated, the equation is

$$N_o = \left(\frac{G_o+B_g N_d}{A_o-Y}\right)^{\frac{1}{R_o+B_r N_d}} - 1 - B_e E \qquad (24)$$

This equation can then be converted to a cost function by using the cost of operational equipment training as the cost base where the cost of one unit of operational training arbitrarily equals one. Device training costs can then be expressed as a proportion of operational equipment training costs. Thus, if *CR* is the cost ratio of operational training cost per unit of instruction divided by device training cost per unit of instruction, and *DR* is the amount of resources provided for device training, then *CR x DR* is the amount of device training that can be supported for the same cost as *DR* operational equipment training units. Thus, substituting *CR x DR* for N_d in the above equation produces

$$N_o = \left(\frac{G_o+B_g CRxDR}{A_o-Y}\right)^{\frac{1}{R_o+B_r CRxDR}} - 1 - B_e E \qquad (25)$$

In terms of the operational equipment cost metric, total cost for training to any criterion is *DR* plus N_o expressed as the above equation:

$$Totalcost = DR + \left(\frac{G_o+B_g CRxDR}{A_o-Y}\right)^{\frac{1}{R_o+B_r CRxDR}} - 1 - B_e E \qquad (26)$$

For any given performance standard, Y, plotting this equation will reveal the amount of device training that minimizes total training costs. Minimum total costs can also be derived by differentiation; however, plotting the data with a spreadsheet program would be much easier and appropriate for the (im)precision of the estimates. Figure 12, shown earlier in the chapter, presents a sample total cost curve showing that minimum total cost occurs when resources equivalent to approximately seven unit of operational training are spent on device training. The actual amount of device training to minimize total cost would be seven times the cost ratio, *CR*. Note that the "dip" in the total cost curve is caused by the fact that the hypothetical device is less expensive to operate than the operational equipment. That is, the tradeoff of device training for operational equipment training is effective because many more training trials can be accomplished on the device than on the actual equipment for the same cost.

Note also that this solution for units of device pretraining is in a sense the "universal solution." That is, it minimizes cost regardless of the amount of extra operational equipment training needed to reach a set standard of performance. Also, it maximizes performance for any given training budget.

Multiple Devices and Mixed Training Orders

The above approach assumes separate equipment training and device training curves and makes the equipment training curve parameters the dependent variables for judging device pretraining. Inherent in these formulations is a fixed order of training: device followed by operational equipment. There are both conceptual (Morrison & Holding, 1990) and practical scheduling reasons for allowing alternating training orders. Unfortunately, this adds another complication to the analysis problem. For example, if training begins on the operational equipment, is switched to a device, and returns to the operational equipment, then under the moderated modeling approach, initial equipment training would moderate device training which in turn would moderate subsequent equipment training. While one could construct another layer of moderating effects in the formulas presented above, it would

be a senseless exercise because we are already straining data requirements. Thus, we propose three alternative compromise models.

Develop a single "task" curve. The above formulations have assumed that learning curves should be constructed separately for each training method. As an alternative we propose using a single *task* learning curve where parameters are a function of the *kind* as well as *amount* of training. The approach also allows analysis of multiple devices mixed with operational equipment training. The parameters, A, G, and R, would be expressed as follows.

Starting point information, carried by G, becomes a function of which kind of training occurs first:

$$G = B_{gd}O_d + B_g \tag{27}$$

where $O_d = 1$ if the device training is used first and 0, otherwise.

Rate and asymptote may be conceptualized as functions of the relative amounts of each kind of training:

$$R = B_{rd}N_d + B_{ro}N_o \tag{28}$$

$$A = B_{ad}N_d + B_{ao}N_o + B_{ado}N_dN_o \tag{29}$$

The interaction term would only be required if an "expanded criterion" were being analyzed and the training methods were expected to train somewhat different aspects of the task. In that case, performance after training on both devices may be higher than an equal amount of training on either method alone.

The combined effects of trial and previous experience would become

$$N + B_eE = B_{eo}N_o + B_{ed}N_d + B_{od}N_oN_d + B_eE. \tag{30}$$

Here, the interaction term is necessary to avoid a linear iso-performance curve.

These new parameter functions could then be substituted in the initial formulas for power and exponential formulas. Because of the number of parameters, choices should be made concerning which curve parameters to analyze as functions. For example, one rationale for alternating between device and operational equipment is to eliminate the "interfacing" adjustment. As a consequence, one might assume that the basic curve shape is constant so that A, G, and R are estimated single parameters, not functions as proposed above, and that the effects of mixed training can be captured by the multi-parameter function for trials above. Thus, given a set of data with varying amounts of training on a device and operational training for which order of training is also mixed, the following power and exponential equations may be used to capture their combined effects:

$$\textit{Power:}\quad Y = A - G(B_{eo}N_o + B_{ed}N_d + B_{od}N_oN_d + B_eE + 1)^{-R} \tag{31}$$

$$\textit{Exponential:}\quad Y = A - Ge^{-R(B_{eo}N_o + B_{ed}N_d + B_{od}N_oN_d + B_eE - 1)} \tag{32}$$

Unfortunately, the equations are complex, each with seven parameters to be estimated. However, they offer the flexibility of including more than one training device by expansion of the equations. For example, the combination of two devices, a and b, and operational training may be captured in the parameters as follows:

$$G = B_{ga}O_a + B_{gb}O_b + B_g \tag{33}$$

$$R = B_{ra}N_a + B_{rb}N_b + B_{ro}N_o \tag{34}$$

$$A = B_{aa}N_a + B_{ab}N_b + B_{ao}N_o + B_{aab}N_aN_b + B_{aao}N_aN_o + B_{abo}N_bN_o \tag{35}$$

$$N + B_eE = B_oN_o + B_{na}N_a + B_{nb}N_b + B_{no}N_oN_a + B_{nb}N_oN_b + B_{ab}N_aN_b + B_eE \tag{36}$$

Again, these parameter functions may be substituted into the basic learning curves. With so many parameters to be estimated, however, solutions to the equations, may be elusive.

<u>Segment the problem by levels of experience</u>. McDaniel, Schmidt, and Hunter (1988) investigated the curvilinear effects of experience on job performance by dividing their cross-sectional data into three groups based on amount of experience. They used standard regression techniques to calculate the slope of the linear relationship between experience and performance within each group. The group with the least experience had the lowest mean performance and steepest slope and the group with the most experience had the highest mean performance and most gradual slope. By combining these results across groups, they assembled an overall experience-performance curve from three straight line segments. This procedure may be adapted to the resource-proficiency tradeoff problem. That is, data may be divided into three groups based on proficiency level. Then, multiple linear regression may be used to examine the relationship between amount of training and proficiency within each group. The regression coefficients for the different training methods will indicate which method, within the given proficiency range, is most effective. Because the analysis is linear, there will be no tradeoff within groups and only one training method can be the optimum. However, different training methods may be optimum for the different groups.

<u>Compare learning curves slopes</u>. The analysis method proposed by Sticha et al. (1990) can be adapted if two conditions can be met. First, it must be assumed that devices do not interact, and that none of the devices interacts with operational equipment training. That is, the "device trials as previous experience" model with no interfacing or synergistic effects applies. Second, either groups-by-trials data, or one-shot transfer data (Figure 5 in Chapter 2) must be available for each alternative device. Then, resource-proficiency relationships can be estimated, adjusted to a common cost metric, and the slopes of these curves compared to determine which training method provided the greatest gain at various levels of proficiency.

Recommendations

In this chapter we have attempted to provide a basic discussion of the problems associated with analyzing training design data in order to create tradeoff functions. The goal of tradeoff analysis cannot be to answer in a simple yes/no fashion if a particular training device or method is more effective than another. There may be no categorical answer to such a question because the answer must be qualified by relative costs and level of training. Instead, the goal is to find the mix of training methods that either maximizes performance for a set training budget, or minimizes cost for a given training standard. It may be that, in spite of cost differences and regardless of level, training is most cost effective if only one particular method is used. On the other hand, overall training may be more effective if a particular portion of training is conducted by one method and the rest by another. Again, the goal of the analysis is to find the optimal use of training methods, including the possibility that only one method should be used.

The basic premise of this chapter is that typical analysis methods that use the general linear model, like ANOVA, cannot capture the complexity inherent in curvilinear tradeoff models. Thus, alternative analytic approaches have been presented. As we warned earlier in the chapter, none of the alternatives are attractive. A large number of parameters is required to describe the combined effects of two training methods. This sort of data analysis is not a cut-and-dry procedure; a certain "spirit of adventure" is needed to fit data analysis techniques to specific research situations.

Chapter 2 presented two basic types of data collection procedures, experimental and nonexperimental. Choosing a data analysis approach is not controlled by these differences in experimental versus nonexperimental control of training resources. Rather, the choice of analysis method hinges on (a) whether or not order of training by different methods is fixed or variable, (b) whether or not the device includes portions of the overall task that are not included in normal equipment training, (c) the number of devices being examined, and (d) the amount of data available for analysis. Figure 17 presents some guidelines for selecting analytic models given these considerations. These are not hard-and-fast rules, only starting points. As we have indicated previously, the analysis may require trial-and-error examination of alternative models. The analysis should also be guided by qualitative observations for the training procedures, particularly if the amount of data is less than adequate.

The sophisticated moderated learning curve models with many parameters can be explored with some degree of confidence only when there are two training methods (e.g., a device and operational training), applied in a fixed order training sequence with lots of data. The weakest position is mixed order training data from several devices and relatively few cases. For this case, the mathematical models must be simplified to as few parameters as possible and even then, the analysis will be subject to the idiosyncracies in the data. For situations more complicated than two devices applied in a fix order, our ability to empirically describe the underlying patterns and relationships is weak. Analysis and conclusions must be augmented by observations and other sources of qualitative data. It may be much more feasible to simplify the problem by investigating training options in pairs.

Design Factor	Options	Recommendation
Number of training options	Two (e.g., device and equipment)	Directly fit iso-performance curve or fit moderated model.
-	Three or more	Fit a "task" curve model, segment the problem by proficiency level, or compare curve slopes. Alternatively, simplify the problem by comparing two options at a time.
Device training includes more of total task than normal equipment training	Yes	Use an expanded criterion measure. Fit a moderated model using an asymptote parameter or use a model that does not have an asymptote, such as log/log.
	No	Omit moderator for asymptote parameter.
Training sequence	Device followed by operational training	Directly fit iso-performance curve or fit moderated model.
	Mixed	Fit a "task" curve model, segment the problem by proficiency level, or compare curve slopes.
Amount of data	Adequate	Use moderated models
	Marginal	Reduce number of parameters as much as possible. Segment the problem by experience.

Figure 17. Analysis guidelines based on research design issues.

It may be possible, if certain assumptions are accepted, to piece together a tradeoff function from data gathered from different sources. That is, if device and operational equipment curves are accepted as fixed so the effect of device pretraining on equipment training can be described as previous experience (i.e., Figure 13), then it may be possible to develop (a) operational curve parameters from one set of data (e.g, Hoffman, 1989), (b) determine device curve parameters for performance on the device on another set of data, and (c) estimate a prediction function relating device performance to operational equipment performance. These functions can then be combined as shown by Cronholm (1985) or as shown by Sticha et al. (1990) to determine the optimum training mix. Although the researcher may be less confident of the validity of the research than traditional confidence levels ascribe, the results may nevertheless provide a useful description of training resource tradeoffs.

Chapter 4. Alternative Data Collection Methods

Implicit in the previous chapters is the assumption that resource-proficiency relationships should be derived from the actual performance of armor crewmen. Although this undoubtedly would be the most valid approach, there are some serious obstacles to obtaining reliable and valid performance data. For one, performance-based research is obviously out of the question for weapons or training systems that are under development.[7] Even for those systems that currently exist, there are fundamental problems related to measuring gunnery performance; these issues are discussed at length by Hoffman, Fotouhi, Meade, and Blacksten (1990) in their review of gunnery performance measurement. The present report has pointed out two additional problems related to the design and analysis of empirical data research: (a) the necessity of intruding into the training process as dictated by experimental designs, and (b) the large sample sizes required to derive valid resource-proficiency relationships. Finally, the process of arranging for and administering the collection of empirical performance consumes considerable time and other resources; research sponsors may not be willing to commit the resources required to answer seemingly simple questions about resource-proficiency relationships.

If at all possible, the researcher should seek to derive resource-proficiency relationships from empirical performance data. However, because of the problems and constraints discussed above, empirical performance-based research is often infeasible. The purpose of the present chapter is to discuss alternative research methods that are appropriate for research-proficiency tradeoff questions.

Research Background

A number of researchers (e.g., Caro, 1977; Hagin, Osborne, Hockenberger, Smith, & Gray, 1981; Pfeiffer & Horey, 1988) have argued that the only viable alternative to using performance data for evaluating training devices is to obtain and analyze informed opinion in a systematic fashion. Pfeiffer and Horey reviewed the aviation research literature for rational analytic approaches to collecting rational data. Although the techniques they identified were diverse in their methodological procedures, Pfeiffer and Horey argued that all have a common reductionistic assumption: that a training system can be analyzed into elemental units from which their benefits may be derived. They classified each of the 18 analytic methods they identified as falling into one of four categories: (a) *index techniques* that evaluate training systems in terms of the presence or absence of predefined attributes, (b) *magnitude* techniques that provide numerical ratings of devices on one or more attributes, (c) *proximity* techniques that involve the comparison of two or more training systems, and (d) *interlocking* techniques that combine ratings of both attributes and systems into a single measure of merit. This taxonomy is useful in describing alternative approaches for evaluating armor gunnery training devices.

---

[7]The exception might be man-in-the-loop, simulation research. However, the functional relationship between performance in the simulation and performance on the yet-to-be-produced equipment remains unknown.

Rational Approaches to Evaluating Gunnery Training Devices

Two general approaches to rational analysis have been used to evaluate gunnery training devices. The first approach comprises a set of related transfer forecast methods developed as a part of a research effort named TRAINVICE (Tufano & Evans, 1982). An example of the magnitude estimation method of rational analysis, the TRAINVICE approach requires subject matter experts (SMEs) to rate a training device in terms of four attributes: (a) the proportion of the task domain covered by the device, (b) the similarity between the controls and displays of the actual equipment and the device, (c) the difference in initial skill/knowledge/performance level and the level required for successful performance on the actual equipment, and (d) the extent to which the device supports effective training practices. The attribute ratings are then combined mathematically to yield an overall score (ranging from 0 to 1) that is intended to describe the device's transfer potential.

Tufano and Evans (1982) characterized TRAINVICE methods as "...the most ambitious steps taken to date in the field of analytic evaluation" (p. 2); nevertheless, empirical research has revealed problems with using TRAINVICE to generate surrogate transfer data for resource-proficiency research. Harris, Ford, Tufano, and Wiggs (1985) examined the reliability and validity of four TRAINVICE methods by having two raters use the methods to rate four different gunnery training devices. Despite fairly good agreement between the two raters in their overall scores for the four devices, the results failed to predict the outcome of an actual transfer experiment. Harris et al. also pointed out that it was not clear how the overall score from TRAINVICE translates to actual performance. Furthermore, except for the attribute related to training difficulty, the TRAINVICE variables are not sensitive to variations in amount of training. This is an essential characteristic for research on resource-proficiency tradeoffs where amount of training is the primary independent variable.

The second approach to rational analysis is an index technique that was used in two evaluations of computer-based gunnery training devices (Hoffman & Morrison, 1988; Campshure, 1990). In these investigations, the researchers used extensive checklists to determine the extent to which the domain of the conditions and behaviors related to gunnery performance were simulated on computer-based gunnery training devices. The results of the analysis were used to prescribe which of the devices were appropriate for training the various objectives related to tank gunnery. One of the strengths of this qualitative analysis is that the results can be used directly to devise training strategies or to propose "fixes" for device design problems. The weakness is that the nominal data obtained from this approach are not appropriate for addressing the quantitative problems of resource-proficiency tradeoffs.

In summary, neither of the two approaches to the rational analysis of gunnery training devices provides data that are suitable for resource tradeoff research. As a result, we decided to examine an analytic method developed outside of the sphere of gunnery research.

An Alternative Technique: Simulated Transfer

A promising alternative analytic method is a magnitude estimation technique that Pfeiffer and Horey (1988) have termed "simulated transfer." Originally developed by Hagin et al. (1981) in their guidebook for evaluating aircrew training devices, simulated transfer is a remarkably straightforward approach to predicting transfer: The rater, an SME familiar with performance both on the to-be-evaluated training device and on the operational equipment, is asked directly to estimate the data that would be obtained from a transfer experiment. As described by Pfeiffer and Horey, the rater generates these data on a task-by-task basis. For each task, the rater is provided an estimate of the number of trials that the average student needs to learn the task to standard in the aircraft. The rater then estimates two values: (a) the number of trials that it would take the average student to learn the task to mastery in the simulator; and (b) given task mastery in the simulator, the number of additional trials that would be required to learn the task to mastery in the actual aircraft. Data obtained from this technique can be used in the same manner as performance data to calculate quantitative indexes such as transfer savings scores and transfer ratios.

The simulated transfer approach offers two important advantages over the two previous analytic techniques developed for gunnery. First, the rated concept (effects of practice on proficiency) is concrete and easily understood by raters, especially military experts such as master gunners who are familiar with performance on the device and the actual equipment. In fact, the rating task is similar to the decisions that SMEs make when laying out training schedules. In contrast, techniques such as TRAINVICE often require the rater to judge more abstract and unfamiliar concepts such as the degree to which the device conforms to training theory. Second, the data resulting from simulated transfer can be treated as is; that is, they do not have to undergo mathematical transformations that would be required by more complex proximity or interlocking techniques. This should speed the process of data analysis and promote quicker responses to user questions about resource-proficiency tradeoffs. It should also make the data more understandable to the user.

The disadvantage to the simulated transfer method as described by Pfeiffer and Horey (1988) is that it provides an estimate of transfer for only one level of device training. As a result, their formulation does not provide two relationships required for tradeoff research: proficiency on the device and on the actual equipment as a function of device practice. Ideally, simulated transfer should be designed to provide data that are similar to those which would be obtained either from a Roscoe-type design or from a groups-by-trials design. Such data would be amenable to the analysis methods described in the previous chapter. Modified to provide this type of data, the simulated transfer method may provide a surrogate research method for determining resource-proficiency tradeoffs in gunnery research.

A Modified Simulated Transfer Approach

In this section we present a modified approach for collecting simulated transfer data. As indicated in the previous chapters, questions about resource-proficiency tradeoffs can differ in their level of complexity. Thus, simulated transfer questions and the data they produce must be adapted to accommodate factors such as the number of training alternatives being considered, the criterion performance of greatest interest, and the costs of

using the various media. Because of these multiple considerations, we have broken the overall problem of devising simulated transfer procedures down into a set of interrelated steps, each of which is described in subsequent sections. The steps include modeling the training situation, formulating the tradeoff question, specifying the performance criteria, defining the dimensions of performance, estimating training cost ratios, specifying appropriate data points, and devising a rating procedures.

To demonstrate the proposed methods, the present section presents an analysis of the Guard Unit Armory Device Full-Crew Interactive Simulation Trainer, Armor (GUARD FIST I) as it was implemented in FY 1991 training by a selected battalion of the Army National Guard (ARNG). This unit provides an exceptionally rich case for analysis because, in addition to GUARD FIST I, a variety of alternative training media (including both computer-based devices and on-tank exercises) were available for training gunnery in FY 1991. The present report details the development of the method used to obtain simulated transfer data and the results of a test of the simulated method.

Model the Training Situation

As suggested in Chapter 2, it is useful to start by developing a model of the training situation to identify the training media and their interrelationships. Figure 18 presents a somewhat simplified model of the gunnery training in the sample ARNG unit for FY 1991. This figure illustrates that gunnery training was divided into two parts: (a) that which occurred during drill weekends from March-June 1991 generally in and around the company armories, and (b) that which occurred during two weeks of annual training (AT) in July 1991 conducted at the unit's major training area.

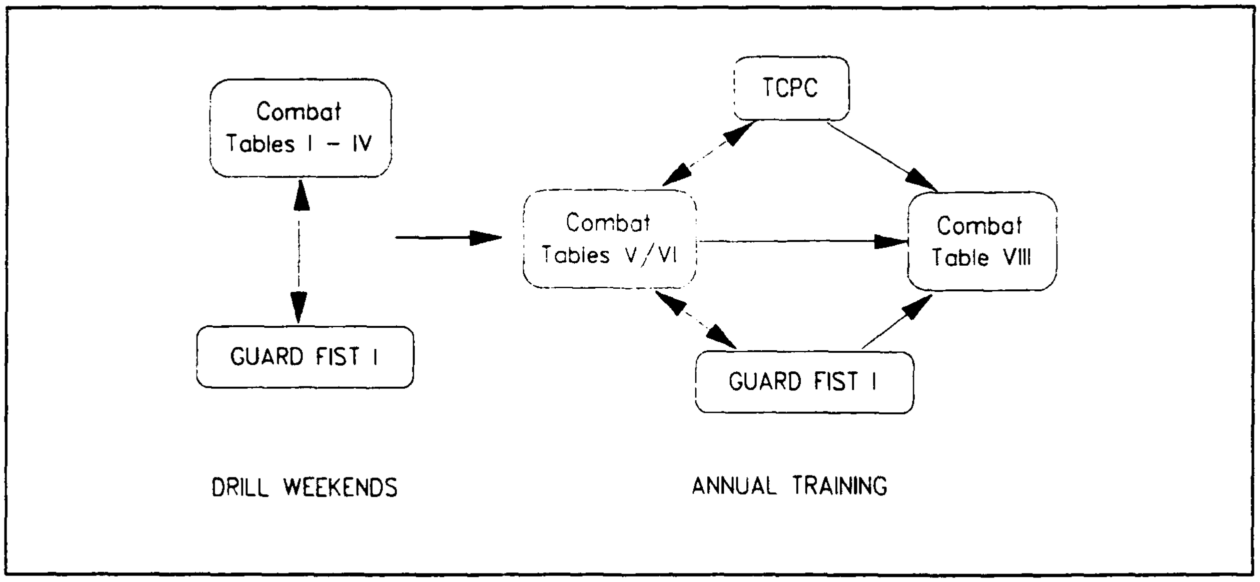

Figure 18. Model of gunnery training media and interrelationships for the selected ARNG unit.

During drill weekends, crew gunnery training was conducted on GUARD FIST I and on Combat Tables I-IV. GUARD FIST I is a computer-based training device that presents computer-generated imagery through the sights of a static tank in power-off mode. The device also enables the tank's controls to be used to engage the computer-generated targets much as they would be used in a live-fire engagement. More details on GUARD FIST I are provided by Morrison, Drucker, and Campshure (1991). Combat Tables I-IV were executed on actual tanks equipped with Multiple Integrated Laser Engagement Simulation (MILES), a line-of-sight laser system that simulates weapon effects.

During AT, crews were trained/tested on two live-fire exercises: (a) a reduced combination of Combat Tables V and VI, and (b) Combat Table VIII.[8] As shown on Figure 18, Table VIII served as the terminal training event in the sequence. In addition, both before and after Combat Table V/VI, crews received training on GUARD FIST I and on the tank. A Tank Crew Proficiency Course (TCPC) was used for training on the tank. The TCPCs consisted of informal dry-fire, target engagement exercises conducted at various locations at the major training area used during AT.

Two other computer-based training devices were used to train gunnery in FY 1991 but they are not represented in Figure 18. One was the arcade-like TopGun device which is designed to provide self-administered training on individual gunner skills. Because of the limited and informal nature of training on this low-fidelity device, TopGun was not expected to have a major impact on crew gunnery performance. The other was the M-COFT which is designed to provide training for critical gunner/TC skills. Work on this device was limited to one or two administrations of a one-hour test in which no extrinsic performance feedback was provided. Ongoing research is being performed to determine whether these tests relate to live-fire crew performance (J. Hagman, personal communication, June 1, 1991). This research will also address the question of whether the tests themselves affected live-fire performance. For the present purposes, however, these two additional devices were ignored.

Formulate the Tradeoff Question

Once the training situation is understood, one can formulate a question about the tradeoff among training resources, which is both important to the user and potentially answerable by these techniques. In the ARNG example, the unit is interested in the amount of time that should be devoted to GUARD FIST I because no guidance currently exists for this prototype device. In the context of the training situation, however, it can be seen that training time on GUARD FIST I is potentially in conflict with training on the tank using either dry-fire techniques alone, dry-fire supplemented with MILES, or live-fire. At the same time, we know that the time available for gunnery training in the ARNG is severely limited. Therefore, an appropriate question might be stated as follows: What mix of GUARD FIST I, dry-fire, and live-fire training

---

[8]Because FY 1991 was a tactical training year, the terminal training event in gunnery would normally have been a Combat Table VII. Additional ammunition was made available to the unit to allow crews to fire Table VIII, the crew qualification exercise. However, crews did not fire a Table VII in preparation for Table VIII, as they would have in a gunnery year.

would yield optimal performance for the limited training time that is available?

Note that this question assumes that live-fire gunnery is treated as just another method for training. This is quite different from traditional ways of thinking about gunnery training. Typically, each of the Combat Tables is treated as a separate training event that each crew must participate in once. If this approach is accepted as a given such that there is no freedom to consider changes to on-tank training, then there is no need for a tradeoff analysis. The addition of a device to the training program is just that--an addition. Without the possibility of reducing or increasing Tank Table training, there is no room for using a training device to create cost or performance optimization. Spending extra training dollars on device training may in fact increase performance. However, performance might be increased even more if a portion of those dollars were devoted to device training and the remainder were spent on additional on-tank training, or alternatively, if greater resources were spent on device training and less spent on training on the tank. It should be noted that such experimental changes to gunnery training would tend to meet considerable resistance, often thwarting efforts to conduct the kind of empirical research needed to answer the tradeoff question posed above. An advantage of simulated transfer research is that it can consider changes to on-tank training without having to manipulate ongoing training schedules.

Specify the Performance Criteria

To go about answering the tradeoff question, we must specify the performance criterion that defines the appropriate outcome of training. For gunnery training, an oft-used criterion is speed and accuracy of live-fire performance on a gunnery range, particularly performance on Combat Table VIII. Hoffman et al. (1990) pointed out that performance measurement on a tank range may be judged deficient because there are combat conditions and circumstances that are not well suited to tank range gunnery. The alternative is to consider a broader concept of gunnery proficiency. A better and more comprehensive measure of gunnery proficiency is combat gunnery; unfortunately, performance data in actual combat is difficult to obtain. Because the simulated transfer approach does not require real performance data, there is no reason to limit the criterion to performance that can feasibly be tested. In other words, a conceptual combat gunnery criterion can be defined and used in simulated transfer research.

In terms of tradeoffs in the ARNG example, the researcher may consider optimizing GUARD FIST I training in combination with tank-range training (a) to maximize Table VIII performance, or (b) to maximize combat gunnery performance. The two solutions may not be the same, particularly if GUARD FIST I trains aspects of combat gunnery that are not trained by tank range gunnery. In this case, more GUARD FIST I training may be recommended if combat gunnery, not Table VIII, were the terminal objective.

The solution adopted was a compromise: A single scale was developed that referenced both combat gunnery performance and performance on the range. To include the tactical component of gunnery, the outcome measure also refers to performance on Tactical Table C (crew qualification), Tactical Table I (platoon qualification), and the platoon exercise from the Army Training and Evaluation Program (ARTEP). The resulting scale is shown in Figure 19. The

scale ranges from 0 for crews who have had no armor training to 100 for crews who would obtain the maximum score on gunnery and tactical exercises and who would devastate the enemy on the battlefield. Using these two end points as anchors, behavioral descriptions were systematically assigned to selected intermediate points on the scale. The resulting scale was constructed such that few, if any, crews would score consistently below 40 or above 90.

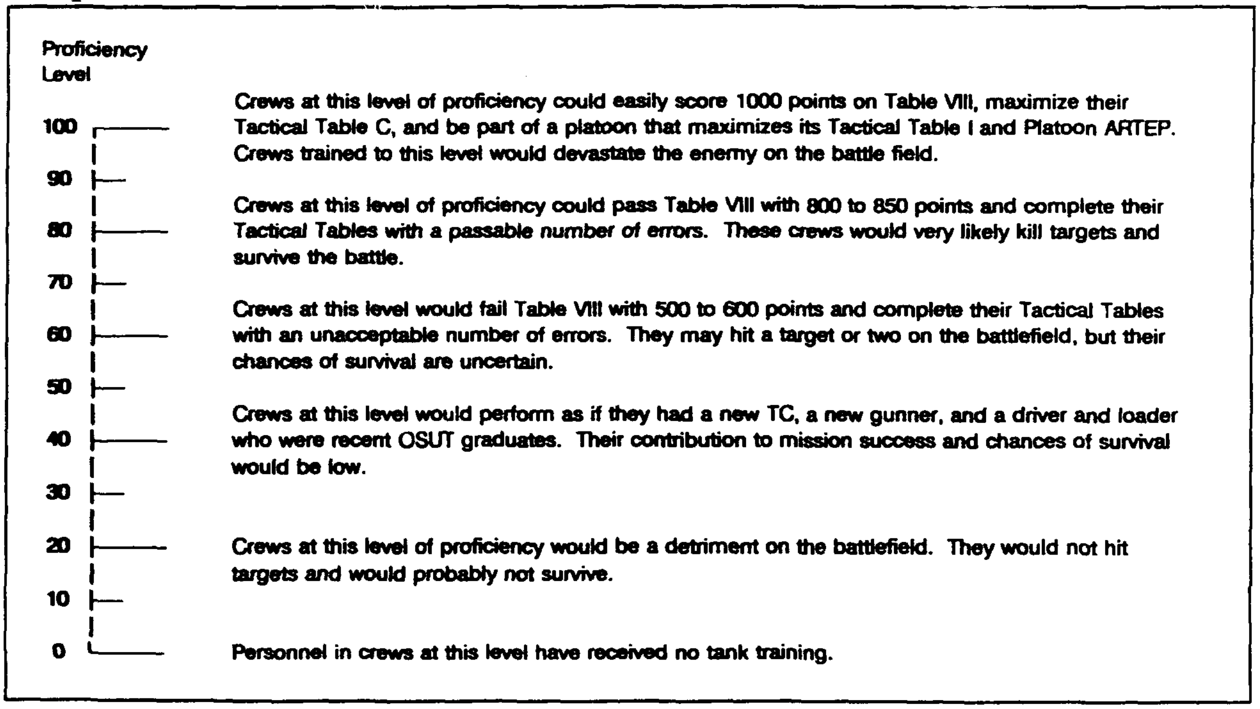

Figure 19. Scale of rating gunnery proficiency.

Define the Dimensions of Performance

Tank gunnery, as well as the performance of most complex skills, is multidimensional in that the criterion performance can be measured in different ways. For instance, Hoffman et al. (1990) distinguished between outcome and process measures of gunnery performance. By outcomes, they meant standard speed and accuracy measures of target hits such as opening times and first round hits. They also argued in favor of the Table VIII system for scoring outcome that combines speed and accuracy of hitting targets with the probability of getting hit by those targets. The latter probability takes into account errors such as staying exposed too long and not engaging targets in order of threat magnitude. Like the previous conceptual rating scheme, the composite outcome score for Table VIII ranges from 0 to 100 for individual engagements, with 1000 points representing maximum score for all 10 engagements in Table VIII.

In contrast to gunnery outcomes, Hoffman et al. (1990) defined gunnery processes as those behaviors (overt or covert) that are crucial to achieving desired outcomes (i.e., target hits and crew survival). These behaviors have been summarized as gunnery subtasks by Morrison, Meade, and Campbell (1990). Figure 20 presents a condensation of the subtasks. A simulated transfer study may consider tradeoffs for each of these components, as Bickley (1980) did for

SEARCH

- Crew searches between and during engagements.
- Crew searches 360°, concentrating in tank's primary sector.
- Crew scans entire sectors/performs detailed searches.

ACQUISITION REPORTS

- Crew transmits brief, timely reports.
- Crew gives accurate target descriptions and locations

CONTACT REPORTS

- Crew immediately reports contact.
- Crew accurately reports direction and target types.
- Crew transmits brief, clear contact reports.

REACTION DRILLS

- Crew's reactions are immediate.
- Crew returns fire on contact.
- Crew turns tank/turret per tactical situation.

MOVEMENT

- Crew uses covered and concealed route or smoke.
- Crew coordinates movement w/ adjacent tanks.
- TC selects appropriate primary, alternate, and supplementary positions.
- Crew properly occupies/moves between hide, turret-down, and hull-down positions.
- TC directs movement out of position to avoid AT fires.
- Driver maintains a steady firing platform and suitable speed.
- Crew avoids untrafficable terrain.

TRACKING AND SWITCHOLOGY

- TC/Gunner use proper ranging/lasing techniques
- TC/Gunner use proper tracking techniques
- TC/Gunner remember to DUMP LEAD after engagement
- TC/Gunner maintain proper switch settings throughout engagement

NORMAL MODE FIRE COMMANDS AND REENGAGEMENT

- TC gives timely, clear fire commands.
- TC selects proper ammo, selects and sequences targets correctly.
- TC gives proper corrections, if required.
- Crew members give timely, correct verbal responses (crew duties).
- Crew reengages missed targets.

DEGRADED MODE AND SUBSEQUENT FIRE COMMANDS

- The crew isolates/corrects/ compensates for degraded conditions ASAP.
- TC gives timely, clear fire commands suitable to degraded condition.
- TC selects proper ammo, selects and sequences targets correctly.
- TC specifies battlesight when appropriate.
- TC gives accurate target descriptions.
- TC gives brief, effective direction element when required.
- TC gives proper corrections, if required.
- Crew members give timely, correct verbal responses (crew duties).
- TC/gunner use standard adjustments/subsequent fire commands per crew or adjacent tank observations.

SPOT REPORTS

- Crew accurately reports threat type, number, and action.
- Crew accurately reports location (within 200 meters).
- Crew accurately reports friendly actions.
- Crew transmits SPOT reports ASAP.

Figure 20. Gunnery components selected for inclusion in test of simulated transfer questionnaire.

his helicopter tasks. Bickley suggested that total cost is the sum of simulator training plus aircraft training for each of the tasks. This proposal implies that training on each task is independent of training on any other task, such that training on any one task on the simulator does not provide any training on the other tasks. The independence assumption does not appear reasonable for gunnery because the components are all intertwined elements of a complex process. Although initial training may isolate practice on the individual elements, terminal performance requires training on the components as an integrated set of activities. In fact, each event included in Figure 18 provides practice on several of the gunnery components. However, the optimum amount of device training for any one gunnery component may not be the optimum amount of training for a related gunnery component. As a result of this interdependence between tasks, combining ratings of separate tasks

into a recommendation for amount of device use could be mathematically complicated.

Estimate Training Cost Ratios

Cost-benefit analyses are plagued by the enormous amount of detail related to training costs. Rather than spending research resources on acquiring extensive costing data, we recommend simply using SME estimates of the relative costs of the training methods under consideration. Obviously, the accuracy of the tradeoff statement depends on the precision of the cost ratio estimates. However, the cost ratios, along with learning and transfer functions presented in Chapter 3, are sufficient to determine whether or not a beneficial tradeoff exists and can be used to provide approximations to the amounts of device training to use.

For most training problems, the highest fidelity training method is usually the most expensive. The highest fidelity method for the ARNG example is live fire which requires significant resources in terms of ammunition, fuel, and soldier time. Relative to live fire, MILES saves ammunition costs because of its laser simulation of weapons effects. MILES training can also save personnel training time related to transporting soldiers to and from ranges. Live-fire ranges are usually located at distant major training areas, whereas the eye-safe MILES laser system can be used in many local training areas nearer the individual armories. Considering savings in both ammunition and transportation, cost ratios of live-fire to MILES training are estimated to be at least 2:1 or 3:1.

When comparing GUARD FIST I to live-fire training, even greater cost ratios may be obtained. No transportation or fuel costs are incurred because GUARD FIST I training can be delivered in the armory using a static tank. In terms of equipment costs, ratios of 5:1 to 6:1 of live-fire to GUARD FIST I costs seem to be a conservative estimation. In addition to equipment costs, an important advantage of GUARD FIST I training is the significant increase in the number of training repetitions that can be achieved compared to training on a range with either live fire or MILES. For example, it may take a crew as much as an hour to move from an assembly area to the tank range start point, traverse the 5 engagements of a Tank Table course, receive a safety inspection, and return to the assembly area for an after-action review of performance; in that same amount of time, a crew could conceivably complete as many as 30 engagements in GUARD FIST I. If time to move tanks to a training area, return them, and perform maintenance is also included in tank-range training time, the GUARD FIST I time advantage is even larger: ratios of 10:1 would not be unreasonable when taking training time into account.

The focus on saving time is particularly important when there is a limit on available training time. For instance, it is possible to arrive at a total cost function specifying a particular number of device exercises and a particular amount of actual equipment exercises that produce a prescribed training level, but that do not fit in the allotted time frame. In addition to constructing a total cost curve, it may also be wise to construct a total time curve using a time ratio instead of a cost ratio. The resulting mix of device and operation equipment training that minimizes total time may not be the same as the mix that minimizes cost, but it may be one that fits within the budget and meets time constraints.

Specify Appropriate Data Points

Bickley (1980) pointed out that specifying amounts of device training is a difficult yet crucial part of tradeoff research. If the values are too large, performance will fall only in the asymptotic part of the learning curve without providing information about the curvilinear portion; if the values are too small, performance will fall only in the curvilinear portion without providing information on the asymptote. If the rate and asymptote of the function can be estimated beforehand, the points may be spaced geometrically or logarithmically to place more points nearer the start than the asymptote of the curve. However, if the shape is unknown, equal intervals between amounts of training is perhaps the better alternative.

An important consideration in choosing data points is the experience level of the target training population. With novice soldiers, the learning curve should be fairly steep--that is, relatively small amounts of training have large effects on performance. With highly experienced soldiers, however, larger amounts of training would be required to significantly affect performance. In addition to affecting the choice of data points, the experience level of the soldiers under consideration must be explicitly stated in the rating instructions.

With regard to the ARNG example, we decided to use equal intervals between increments of engagements or trials because little was known about the shape of the acquisition curve. Increments or blocks of 10 engagements were used to match the number of engagements in a typical Combat Table. With regard to experience, we decided to specify the novice level of experience for two reasons. First, we wanted to obtain data for which training effects should be pronounced. Second, we felt that crews with high levels of experience were a rare occurrence, and that the concept of the "average" crew was too ambiguous to rate.

Devise Rating Procedures

The next step in devising a simulated transfer procedure is to conceive, in a general fashion, the overall rating procedures. The objective here is to generate data that can be used to construct equations for tradeoff calculations. Depending on the number of training alternatives under consideration, rating questions may be constructed that generate the following types of data:

1. Groups-by-trials data: Raters would enter values indicating performance levels in matrices patterned after the groups-by-trials design (See Figure 8 or Variation 3 in Figure 9 in Chapter 2).

2. Iso-performance data: For differing amounts of device training, raters would indicate the amount of operational equipment training required to reach a selected performance standard.

3. Learning curve data: Raters would indicate proficiency expected after selected amounts of training. Ratings would be gathered separately for each training alternative and would be rated in terms of the same terminal performance criterion.

The first and second options are most appropriate for tradeoff problems that include only two training alternatives presented in a specified training order. Data from the first option can be analyzed using the moderated analysis methods described in Chapter 3 to generate iso-performance and total cost curves in order to find the amount of device training that maximizes performance or minimizes cost. In contrast, data from the second option provides iso-performance curves directly, relieving the researcher from having to make the mathematical modeling decisions described in the previous chapter. For two training alternatives, the number of required judgments is manageable for these options; but for more comparisons, these options require many judgments.

For estimating the tradeoffs among three or more training media, the third option appears most reasonable. To use this option, however, the researcher must be willing to forego analysis of potential moderator relationships related to asymptote, curve start point, or learning rate. It assumes that transfer from one device to another or to the operational equipment is realized immediately without any "interfacing" decrement. It should be pointed out that the interfacing decrement is a performance phenomenon, not a proficiency phenomenon. That is, while initial performance may suffer, the underlying knowledges and skills learned on the device remain, allowing performance on the equipment to improve rapidly. This assumption is less of a problem for simulated transfer than for an empirical study because the raters can be directed to focus on underlying proficiency. The learning curves that result from this approach, adjusted by cost ratios per Chapter 3, can be compared in the manner suggested by Sticha, Schlager, Buede, Epstein, and Blacksten (1990): For a given level of proficiency, the training alternative with the steepest learning slope is the preferred alternative.

The third option was the one chosen for the analysis of GUARD FIST I. The raters, who were platoon sergeants and master gunners that had administered training on GUARD FIST I, were asked to judge the expected proficiency of crews after 0, 10, 20, 30, 40, and 50 engagements on GUARD FIST I as well as MILES and live-fire training. As discussed earlier, proficiency was judged only on those gunnery subtasks that were trainable on each medium. In addition, the raters were asked to rate the confidence that they had for rating subtask proficiency on each medium. The resulting rating form, the Training Prediction Questionnaire, with all instructions is provided in Appendix A.

Increase Reliability of Ratings

To obtain valid ratings, SMEs should be carefully chosen to have knowledge of performance on both the training device and the actual equipment. SMEs should be unit soldiers who have recent experience conducting gunnery training on the devices under consideration. For the sample ARNG unit, the training NCOs were chosen as raters. In armor units, these NCOs are usually master gunners and therefore have substantial technical knowledge of gunnery. In conjunction with their commanders, the training NCOs plan for and conduct all gunnery training. Thus, they are familiar with the effects of all three media on gunnery proficiency.

The fact that SMEs are knowledgeable about the problem does not, however, mean that they will agree on their estimates. There are several elaborations or interventions to the rating procedures that could potentially reduce differences between raters, thereby increasing interrater reliability.

One elaboration on the rating procedure is to require SMEs to review their own responses, but from a different perspective. To this point, SMEs have estimated proficiency at different points in training. The researcher should plot these estimates and let the SME review a graph of the resulting function. This feedback emphasizes the *pattern* of SME estimates and may reveal inconsistencies in their responses. Upon review, the raters should then be allowed to revise their estimates accordingly. Providing such graphic feedback potentially increases reliability by eliminating inconsistencies within individuals.

Another intervention for increasing reliability is to use consensus-building techniques, such as the Delphi method, to reduce differences between SMEs. The Delphi technique assumes that a group's consensus estimate, achieved after several rounds of anonymous judgments, is more accurate than individual judgments. This technique could effectively be combined with graphic feedback with the stipulation that the consensus building occurs *after* providing graphic feedback to individuals. Having seen and adjusted the pattern of their own responses would make SMEs more aware of the differences in the pattern of other raters' estimates. This should enable the raters to discuss differences at the level of a learning function rather than at the level of individual data points.

The problem with these elaborations on the basic procedure is that they are quite time consuming. This posed a particular problem for the ARNG project. The NCOs provided their ratings as they completed Table VIII at AT. They had little additional time to elaborate on their responses. Another problem related to consensus building is that training NCOs in an armor battalion are not centrally located; their offices are in their respective company armories, which are distributed throughout the state. Because of these difficulties, these elaborations were not executed for the ARNG project.

Results From Tryout of the Simulated Transfer Method

Because the proposed method is experimental, data could not be accepted at face value. Four criteria were identified to judge the success of the method. First, responses from the NCOs should be reliable, showing interrater agreement. Second, ratings across increasing amounts of practice should conform to learning theory expectations by showing some tendency toward a monotonically increasing but negatively accelerating learning curve. Third, ratings should differentiate the methods for teaching the different gunnery components in ways that reasonably conform to expectations from previous reviews of the methods (e.g., Morrison, Campshure, & Doyle, 1991; Hoffman et al., 1990) and conventional wisdom. Finally, results should show agreement with empirical data from the ARNG collected during the same time period.

The Training Prediction Questionnaire was administered during ARNG annual training to nine NCOs who had acted as instructor/operators (I/Os) for GUARD FIST I training. In addition, one of the four company commanders accepted an invitation to complete the questionnaire. Two to the four companies used only a single I/O for all of their GUARD FIST I training and consequently provided only one rater each. The other two companies used multiple I/Os, and each provided four raters. For three companies, the questionnaire was administered after all gunnery training events, including Table VIII. For the remaining company, time constraints led to the

questionnaire being administered the day before Table VIII, but after all prior gunnery training.

Interrater Reliability

Interrater reliability for each combination of gunnery component and training method was computed using generalizability theory by partitioning ratings according to rater variance, training amount variance, and the interaction the two. These variance estimates were then appropriately combined into reliability coefficients that treat training amounts as fixed, raters as random, and include both rater variance and rater-by-training amount variance as measurement error. Formulas may be found in Brennan (1983).

Table 4, on the following page, presents the reliability estimates for mean ratings from the ten raters. They are uniformly high, ranging from .92 to .97. Those reliabilities are elevated by the strong demand characteristics of the instrument to give successively higher ratings for increasing amounts of training. More important than the reliability *per se* is the standard error of measure that can be used to interpret confidence intervals around each mean proficiency estimate. Standard errors of measure, also presented in Table 4, range from 2.6 to 4.5. Thus, 95% confidence interval around the proficiency estimates range from ±5.1 to ±8.8.

Estimated Learning Curves

Figures 21-23 present estimated learning curves based on raters' mean estimates of proficiency for successive amounts of training and for the expected maximum proficiency or asymptote. The "curves" obviously have a linear tendency. Indeed when the mean proficiency ratings are regressed onto the six levels of training (0 task to 50 tasks), Pearson linear correlations are all over .99. Furthermore, the mean asymptote ratings appear to be straight line projections as if they were ratings of proficiency for the next equal interval in the 0 to 50 series, i.e., 60 tasks. In other words, expected the nonlinear patterns did not emerge.

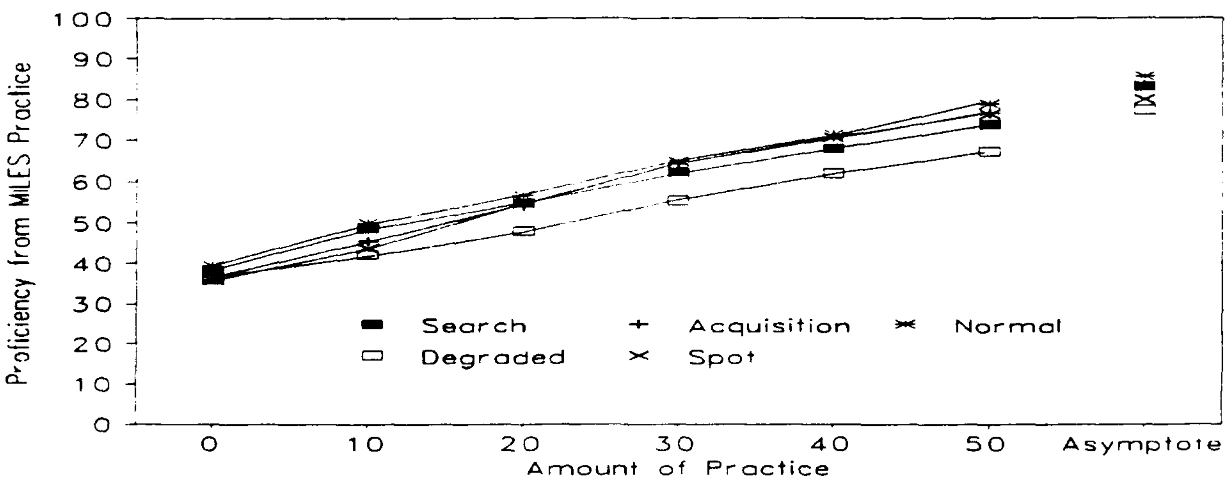

Figure 21. Estimated learning functions for MILES training.

Table 4

Interrater Reliability for the Training Prediction Questionnaire

Training Method Rated Components	Reliability	Standard Error of Measurement
MILES		
Search	0.96	2.76
Acquisition Reports	0.95	3.51
Normal Mode Fire Commands	0.95	3.15
Degrade Mode Commands	0.95	2.66
Spot Reports	0.93	4.29
GUARD FIST I		
Search	0.92	3.99
Acquisition Reports	0.95	3.24
Reaction Drills	0.97	2.56
Normal Mode Fire Commands	0.96	3.23
Degraded Mode Commands	0.97	2.69
Tracking/Switchology	0.93	3.77
Spot Reports	0.92	4.54
Live Fire		
Search	0.97	3.04
Acquisition Reports	0.90	4.35
Normal Mode Fire Commands	0.95	3.51
Degraded Mode Commands	0.96	3.17
Tracking/Switchology	0.97	2.99

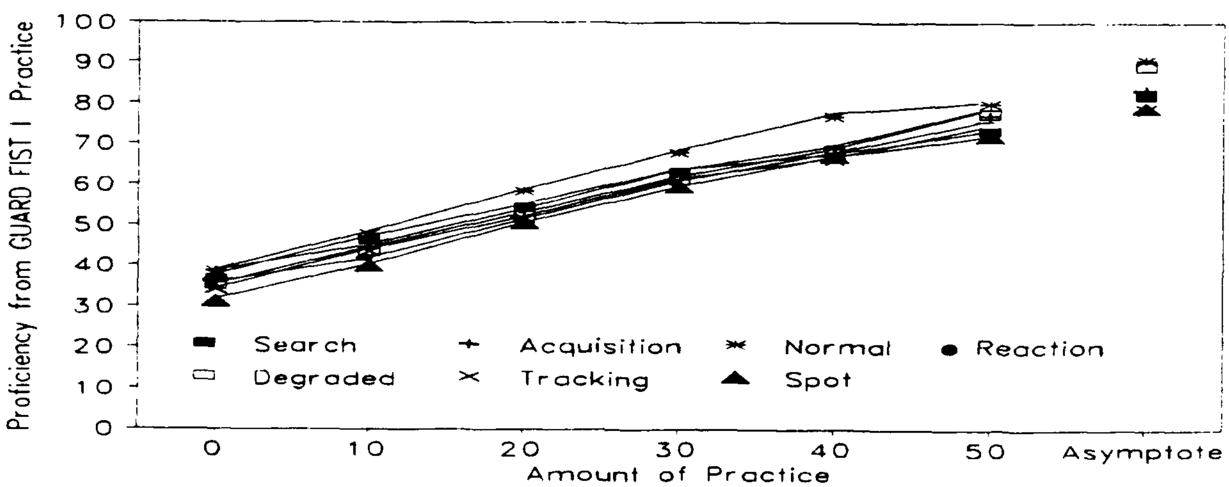

Figure 22. Estimated learning functions for GUARD FIST I training.

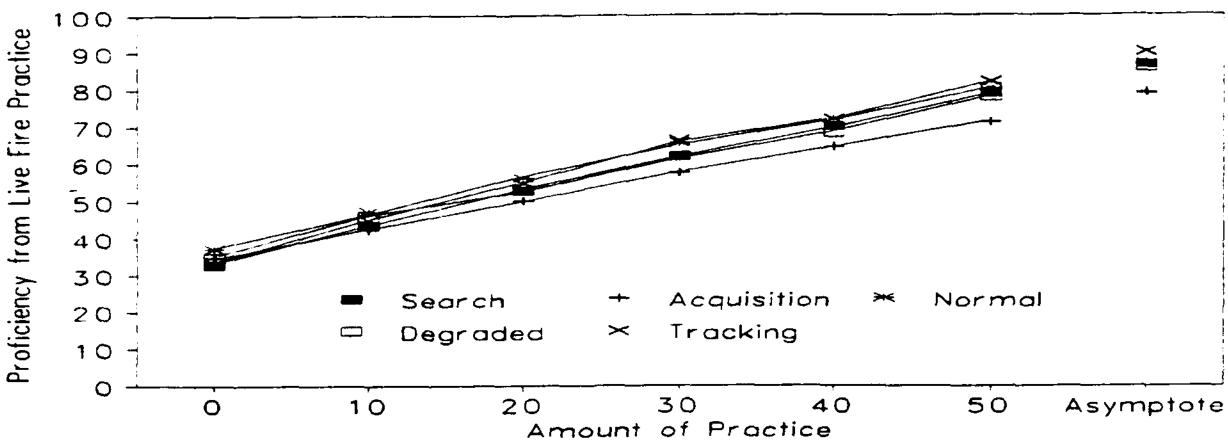

Figure 23. Estimated learning functions for live-fire training.

Figures 21-23 are arranged to reveal any differences in training effectiveness between gunnery components within each of the three training methods. Repeated measures analysis of variance (ANOVA) were conducted for each training method using amount of training (0 to 50 tasks) and gunnery component as repeated factors. An expected and rather trivial result was the highly significant effects for amount of training (*p*s < .001 in all three cases). Only MILES training, however, showed significant differences between the gunnery components in mean expected proficiency for training trials 0 through 50, F (4,36) = 3.48, $p < .05$. Degraded mode training appears lower than the other components. Live-fire training did show a significant amount-by-trials interaction for expected proficiency, F (20,180) = 1.68, $p < .05$. This result appears to stem from the lower learning slope for target acquisition.

Asymptote differences between gunnery components were also tested for each training method. Curiously, GUARD FIST I was the only method to show significant differences in expected asymptote for the different gunnery components, F (6,54) = 2.51, $p < .05$. Judging from the asymptote results alone, GUARD FIST I's strengths appear to be with normal and degraded mode fire commands. With a relaxed alpha level, asymptote differences in amount the gunnery components also appeared for live-fire training (F [4,36] = 2.19, $p < .10$), with target acquisition lower than the other components.

Figures 24-29 display differences between methods in training effectiveness for the different gunnery components. These differences were tested using repeated measures ANOVAs on each of the six gunnery components using amount of training (0 to 50 tasks) and method of training as repeated factors. Trials effects are again significant for all gunnery components (*p*s < .001). No methods effects were found, but in two cases significant methods by trials interactions were obtained. These occurred for degraded mode gunnery (F [10,90] = 2.14, $p < .05$) and for tracking and switchology (F [10,90] = 3.84, $p < .01$). For degraded mode gunnery, MILES training appears less efficient per training trial than GUARD FIST I or live-fire training. For tracking and switchology, GUARD FIST I appears less efficient per training trial than live fire.

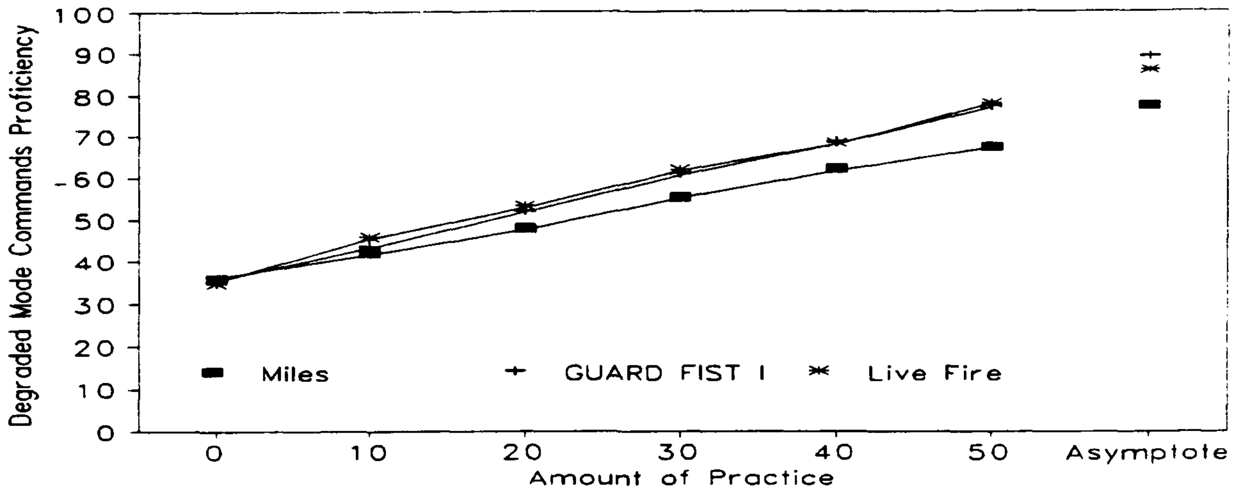

Figure 24. Estimated learning functions for degraded mode fire commands.

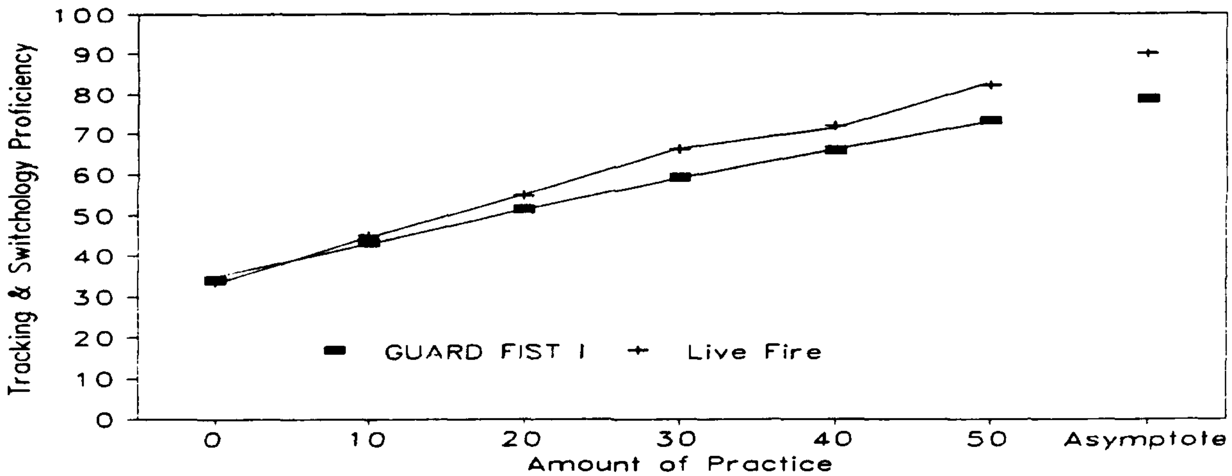

Figure 25. Estimated learning functions for tracking and switchology.

Repeated measures ANOVAs were also conducted for each gunnery component on asymptote estimate. None of the asymptote differences were significant at the traditional $p < .05$ level. However, at the $p < .10$ level, GUARD FIST I and live fire differed in their tracking and switchology, $F\ (1,9) = 4.00$. asymptotes with GUARD FIST I having the lower asymptote.

Conclusions

Three criteria were given for judging the success of the simulated transfer methodology including interrater agreement, conformance to learning theory, and conformance to previous analysis. Results are mixed, at best, indicating that the simulated transfer method as executed does not provide a definitive assessment of the training methods.

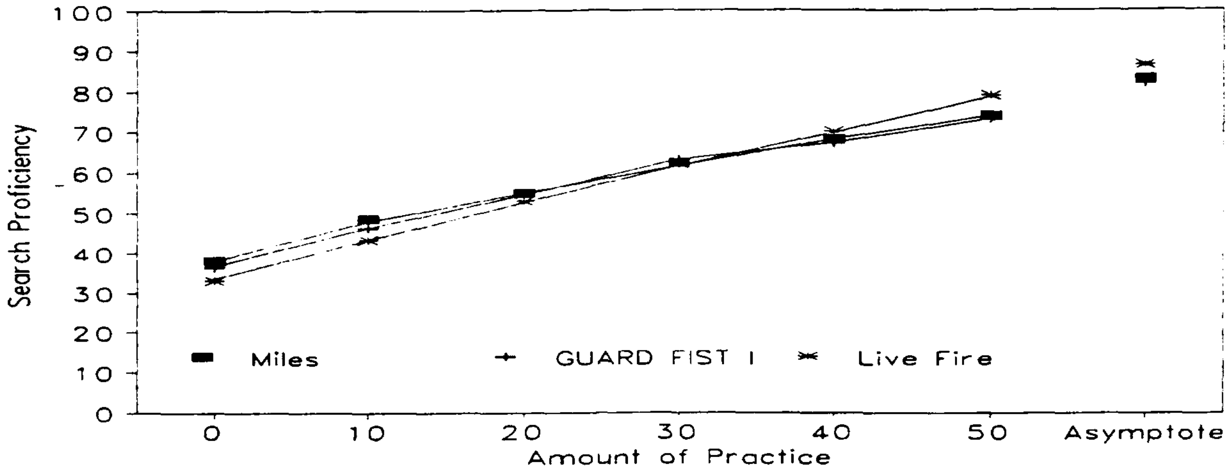

Figure 26. Estimated learning functions for search.

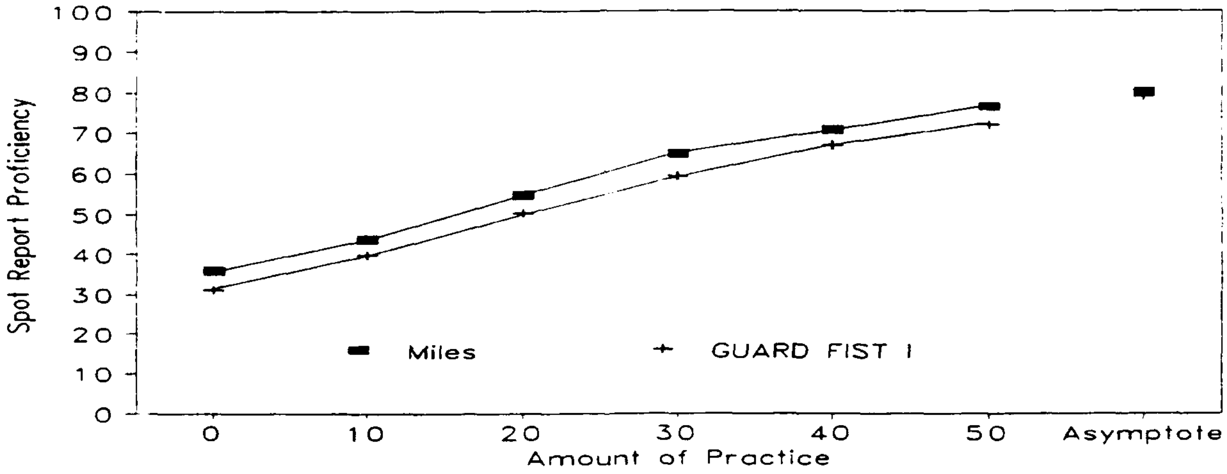

Figure 27. Estimated learning functions for spot reports.

With regard to interrater agreement, the reliability coefficients appear large enough, but they are somewhat deceiving. Reliability is a function of both total variance and error variance. Because proficiency ratings increased systematically across trials, there is sufficient total variance to inflate the coefficients in terms of the typical frame of reference for interpreting reliability. Standard errors of measurement, on the other hand, are not influenced by total variance. With confidence intervals based on these standard errors of measurement (ranging up to ±9) being in the same neighborhood as the differences in estimated proficiency between successive amounts of training, the low rate of significant differences between components within methods and between methods within components is not surprising.

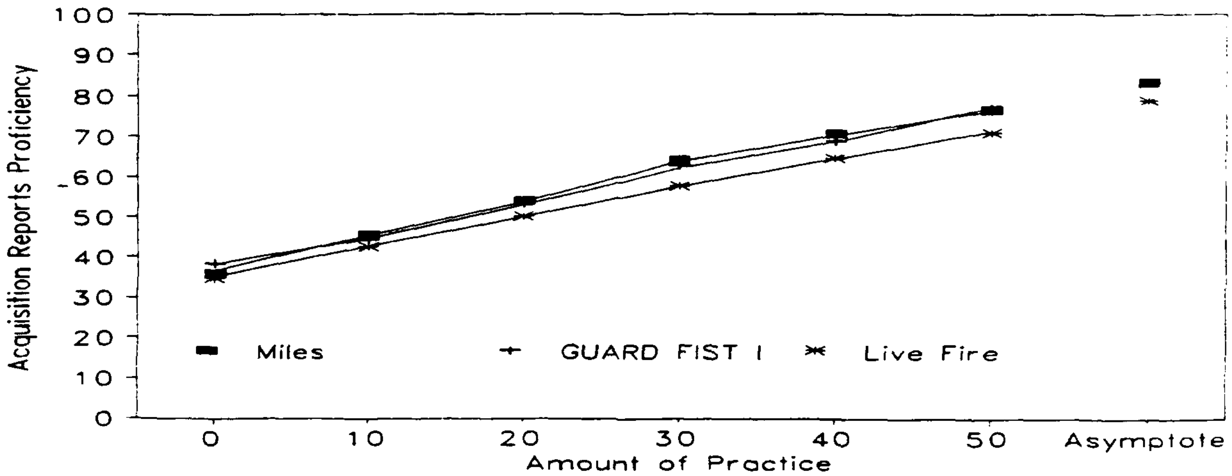

Figure 28. Estimated learning functions for acquisition reports.

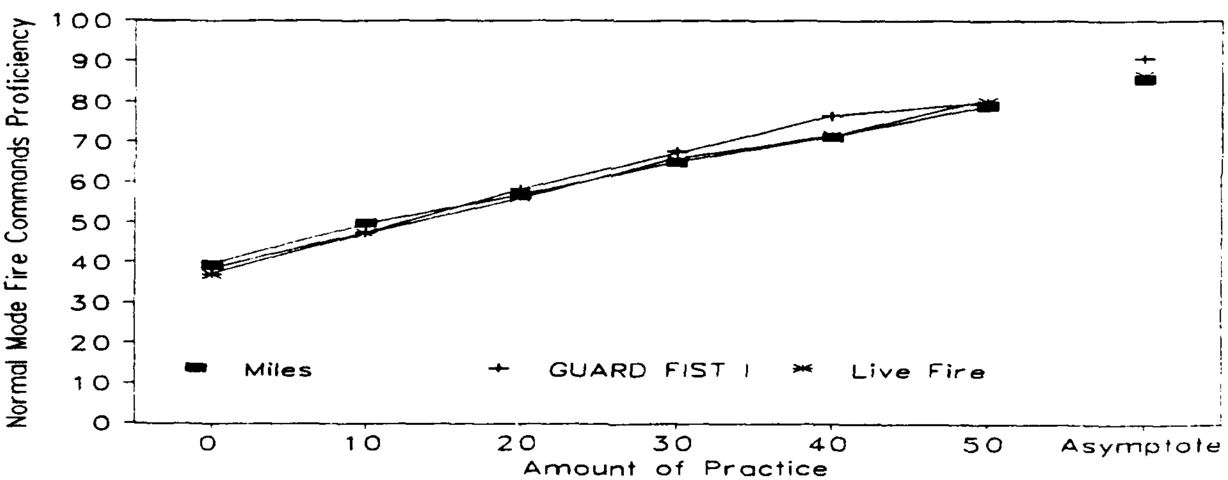

Figure 29. Estimated learning functions for normal mode fire commands.

Our expectation that the estimated learning functions would have some curvilinear tendency was totally unconfirmed. Indeed, it appears that SMEs were simply giving equal interval responses. Although it might be suggested that we queried too small of a range of training trials for a curve to appear, that argument is nullified by the asymptote data. Asymptote responses appeared to be simply another equal interval step beyond the 50-trial responses.

In terms of the congruence of the results with previous analysis and conventional wisdom, there are two observations. First, the differences that did appear between training methods and components were predictable and there were no outstanding differences that failed to appear. Thus, the questionnaire picked up the gross differences we knew in advance and did not

help to fine tune our knowledge of the training methods. Second, SMEs appear to have underestimated how long it takes to acquire the gunnery skills. Based on the scale anchors, one would expect a crew to pass Table VIII if they were rated at about 70% proficiency. Looking across training methods and components, a proficiency of 70% would be expected of a novice crew after they have fired only about 40 engagements. This expectation does not match the battalion's training history and 1991 Table VIII pass rate. The average crew fired more than 40 engagements, counting all of the different gunnery training events prior to Table VIII. For example, the M-COFT exercise itself contained 40 engagements; MILES exercises on Table II, III and IV contained 30 engagements; and many of the crews approached 40 engagements on GUARD FIST I. The four companies ranged from approximately 60 to 130 total engagements prior to Table VIII.

Given these results, it seems unwise to draw any hard-and-fast conclusions about the training methods. However, despite the problems with the simulated transfer data, the general lack of differentiation between the methods compared to the vast difference in the amount of time it takes to train on the methods supports allocating a sizable portion of training to GUARD FIST I. The data are not trustworthy enough to specify an amount of GUARD FIST I training time, but it does appear that there is a time advantage to using GUARD FIST I.

Recommendations

To reiterate the arguments provided at the beginning of this chapter, the most valid approach to deriving resource-proficiency relationships is to use actual performance data. Within the context of simulated transfer, a number of alternative procedures have been proposed. Our initial attempt with a relatively simple version of the method met with limited success at best.

For the method to remain viable, the best recommendation is to perform further research to refine these procedures. Research should be performed to compare and contrast results from the alternative formats proposed for gathering simulated transfer data. Certainly iterative feedback of results to SME showing the implications of their ratings should be given. Comparisons may indicate some of the implicit assumptions that SMEs make about the process of transfer. Other reasonable questions are whether graphic feedback aids the raters to provide more reasonable (i.e., curvilinear) functions, and how quickly consensus can be reached on the shape and location of learning curves. Note, however, that extensive interventions in the data collection process begin to smack of data manipulation to reach preordained conclusions.

Chapter 5. A Test of Nonexperimental Case-Study Methods: Effects of Training on GUARD FIST I and on the Tank

The overall purpose of the research described in this chapter is to test the viability of using the correlational, case-study method to determine tradeoffs concerning two tank gunnery training resources: the Guard Unit Armory Training Device, Armor (GUARD FIST I) and training on the tank itself. A secondary objective of the research was to generate performance data that could be used to provide recommendations for allotting time to GUARD FIST I training.

As asserted in Chapter 2, a case-study can provide two types of data: quantitative and qualitative. The quantitative data are required to determine resource-proficiency relationships. Unfortunately, the quantitative data obtained from case-studies are hampered in two ways: (a) the case study usually provides only a limited number of cases, and (b) the relationships between resources and proficiency are correlational in nature. On the other hand, the strength of the case study is that it provides the opportunity to qualitatively study training in depth and in a realistic context. These qualitative data can be used to comment on the efficiency and effectiveness of training procedures. The present chapter presents an analysis and discussion of both types of information: The first section of the present chapter presents an analysis of the quantitative data obtained from this case study, and the second presents a discussion of the qualitative observations that we made while collecting those data.

Analysis of the Quantitative Data

Method

The general approach was to observe crews in a single Army National Guard (ARNG) armor battalion as they underwent gunnery training and prepared for Combat Table VIII during their two-week annual training (AT). No attempt was made to influence the type or amount of training in the unit. The only constraint imposed on the unit was to record the gunnery training that occurred during the year for each crew.

Experimental participants. Data for this research were obtained from tanks crews assigned to the four companies in the ARNG armor battalion. In keeping with unit training practices, a tank crew was defined by tank commander (TC) and gunner (GNR) without regard to the other two tank crewmembers, the driver (DVR) and loader (LDR). This practice was used for two reasons: (a) TCs and GNRs were trained together to foster appropriate interactive skills between these two key crewmembers; and (b) the battalion did not have a full complement of drivers and loaders for each TC/GNR combination. For gunnery training requiring all four crewmembers, DVRs and LDRs were often temporarily assigned to crews on an as-needed basis. Even with this practice of "hot-seating" DVRs and LDRs to fill out crews, the companies did not have enough TC/GNR pairs to man their full complement of 14 tank crews. Also, the actual number of crews available for training varied from drill week to drill week.

Training events. Two systematic methods were used to train gunnery in this unit in FY 1991.[9] One was training on the tank itself, which was executed in accord with the series of on-tank exercises described in Tank Combat Tables M1, FM 17-12 (Department of the Army, 1990). The second was training on GUARD FIST I. GUARD FIST I is a computer-based, tank-appended device that is designed to be used in an armory setting. (For a more detailed description of GUARD FIST I and its capabilities, see the review of gunnery ARNG training devices by Morrison, Drucker, and Campshure [1990].) GUARD FIST I did not become available for training until late in the training year (four months before AT). Because training time was limited, the battalion did not follow the GUARD FIST I training strategy outlined in the device's technical manual (Daedalean, 1990). Rather, the limited training time was devoted to practicing those GUARD FIST I exercises that mirrored ongoing training on the Combat Tables.

Gunnery training can be divided into three phases. The first phase was the four-month period prior to AT where gunnery training was carried out in and around the local armories. During this phase, training focused on Combat Tables I-IV. Most of the training was conducted on the tank using the Multiple Integrated Laser Engagement Simulation (MILES) system to simulate weapons effects. Although each company had access to MILES, the actual conduct of training with this device varied considerably from company to company depending on local resources and constraints. For instance, one of the units used the adjacent fairgrounds which permitted limited movement of tanks, whereas another unit used the nearby airport which permitted practice on stationary engagements only. In contrast, a third unit was located close to the battalion's major training area and was able to use the considerable range resources available there. With respect to GUARD FIST I, only 3 of the 4 companies were sufficiently close to the site where the device was located to use it for training prior to AT.

The second phase of training occurred during AT prior to Table VIII qualification. AT was conducted in the unit's major training area. Training on the tank comprised both dry-fire exercises, collectively referred to as the Tank Crew Proficiency Course (TCPC), and a preliminary live-fire exercise which combined selected engagements from Combat Tables V and VI. In addition, the GUARD FIST I system was moved to the major training area to avail crews training on this method as well. The battalion scheduled all companies for an initial GUARD FIST I session followed by at least one TCPC prior to Table V/VI. Additional TCPCs and GUARD FIST I sessions were scheduled given time and availability. Although the battalion scheduled and coordinated training during this phase, the actual execution and evaluation of training was left to the individual companies.

The third and final phase of training was the crew qualification exercise, Table VIII. Although this event occurred during AT, it differed from previous events in two ways. First, the purpose of Table VIII was primarily evaluation, whereas the purpose of previous events was primarily training. Second, Table VIII was conducted by permanent-party range soldiers and scored by external evaluators from the unit's round-out division.

---

[9] As noted in the previous chapter, M-COFT and TopGun were also used. M-COFT, however, served only as a testing medium. TopGun was used individually on an informal basis.

Data collection instruments. To provide a common metric for measuring performance on the wide variety of training events, a Training Event Inventory was constructed. Those soldiers who were directly in charge of training crews were asked to fill out a Training Event Inventory on every crew for every training event. If the unit executed both day and night (A and B) portions of a Table, the two portions were considered as separate training events, and forms were filled out for both. The Training Event Inventory is a 2-page document, presented in Appendix B, that directed the trainer to rate each crew's performance on 9 components of gunnery performance (e.g., search, movement, normal mode fire commands and reengagement). The rationale for this approach to scoring gunnery behaviors is explained by Hoffman, Fotouhi, Meade, and Blacksten (1990) in their review of gunnery performance measurement. Some of the components were not trained in some of the exercises. For instance, movement could not be rated on Table II which exclusively consists of stationary engagements. As a result, the raters had the option of responding "not applicable" (NA) to components that they felt were not trained in a particular event.

Results

A preliminary analysis of component differences revealed no important differences among the ratings of the components. That is, the raters appeared to rate the crews similarly over all components. Because of these results and the missing data caused by NA responses, the individual ratings for a crew were combined by computing the average for the rated components, that is, not including those that received an NA response.

Validity of ratings. To examine the validity of the proficiency ratings, the last mean ratings on GUARD FIST I and the last on-tank exercise prior to Table VIII (Table V/VI) were correlated with the scores received on the qualification table. These results, shown in Table 5, indicate that there were positive correlations between the ratings on GUARD FIST I and both day and night versions of Table V/VI. Although neither correlation reached conventional values of significance, the correlation between ratings on GUARD FIST I and Table V/VI (night) approached significance at the .05 level ($p < .08$). There were also positive correlations between the rating on Table V/VI and the scores on Table VIII with the correlation between the rating on the night portion of Table V/VI and the day portion of Table VIII reached significance at the .05 level. In contrast, the correlations between GUARD FIST I and Table VIII were practically zero. This lack of correlations may be explained by the fact that, in most cases, there were one or more intervening training events between the last GUARD FIST I rating and Table VIII.

Overall, results showed a pattern of positive correlations between ratings on training events and Table VIII. Although the correlations between adjacent events were small, they were at least as large or larger than correlations within events--that is, between day and night versions of Table V/VI ($r = .22$) or between day and night versions of Table VIII ($r = .23$). These findings support the contention that the ratings were valid indicators of gunnery proficiency, at least in the short term.

Profiles of training performance. Profiles of mean performance were constructed for the training events in the chronological order that they occurred. The purpose of these profiles was to detect systematic changes

Table 5

Correlation of Ratings on GUARD FIST and Table V/VI with Scores on Table VIII

Variables	1	2	3	4	5
1. Last Rating on GUARD FIST I	1.00 (30)				
2. Rating on Table V/VI (Day)	.25 (30)	1.00 (43)			
3. Rating on Table V/VI (Night)	.34 (28)	.22 (41)	1.00 (43)		
4. Table VIII Score (Day)	.05 (29)	.12 (41)	.31* (41)	1.00 (45)	
5. Table VIII Score (Night)	.10 (28)	.14 (40)	.14 (40)	.23 (42)	1.00 (44)

Note. Sample sizes for correlations are provided in parentheses.

* $p < .05$

in performance that occurred over time. Because of the differences in training among companies, these profiles were constructed separately for the four companies in the battalion. The companies were arbitrarily numbered 1-4. These profiles, shown as Figures 30-33, are split according to the three phases of training: the training that occurred prior to AT, the training that occurred during AT, and training on Table VIII.

An examination of the order of events along the horizontal axis shows that, except for the final Table VIII test, units were trained on a variety of events and in different orders. This complicates the matter of sorting out the effects of the two primary training methods. Focusing on mean performance over those events also reveals no clear upward or downward trends in performance. The exceptions to this generalization were the lower ratings that crews received on the initial GUARD FIST I session during AT and on the final Table VIII. There are at least two explanations for the lack of clear trends in performance. The first is that the proficiency ratings were specific to particular training methods; for instance, a crew performing well on GUARD FIST I may not necessarily perform well on the tank. Although plausible, this explanation is not supported by the present data: Except for Table VIII, none of the training methods consistently produced high or low performance ratings. The second explanation was the fact that different raters were used among and within events. This explanation helps explain the lower ratings on Table VIII. For this event, external evaluators were used as opposed to company soldiers, who rated all other events. It is reasonable to

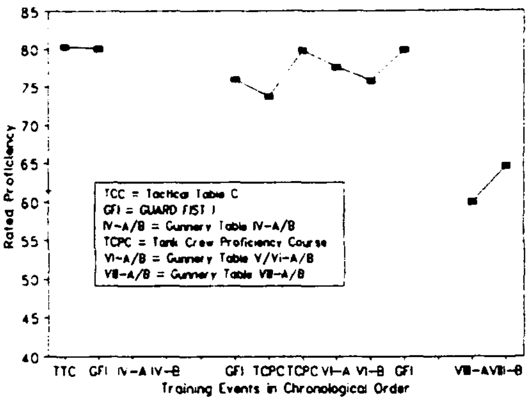

Figure 30. Training performance profile for Company 1. (Data from Table IV are missing.)

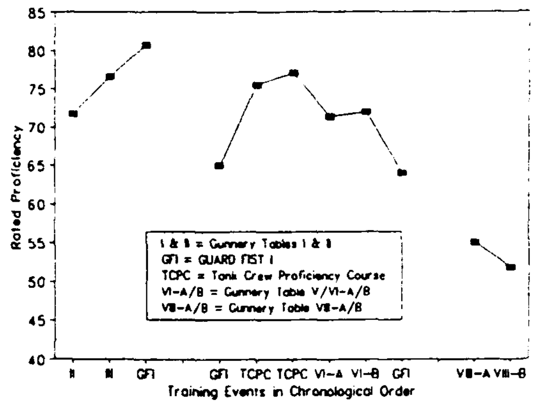

Figure 31. Training performance profile for Company 2.

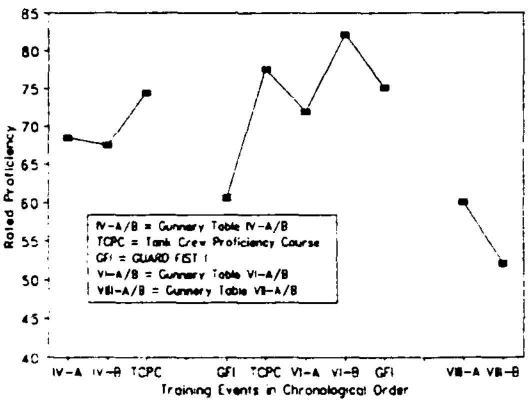

Figure 32. Training performance profile for Company 3.

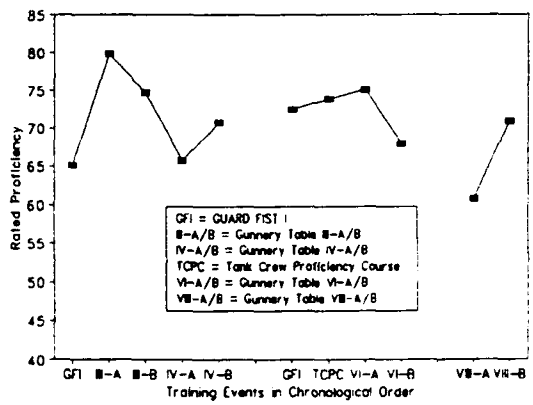

Figure 33. Training performance profile for Company 4.

expect that external evaluators would rate companies lower than soldiers within the company.

Data in the previous section showed that ratings of crews were reasonably stable across adjacent training events and consistent with Table VIII scores. However, the present section showed large differences in mean performance from event to event. These findings suggest a "calibration" problem is inherent in the ratings. The calibration problem does not invalidate the ratings for judging relative proficiency (i.e., the rank ordering) among the crews. However, the calibration problem does not allow the tracking of performance changes of crews from event to event, which was the primary reason for using the proficiency rating method.

<u>Within-crew changes in performance</u>. Part of the large differences in mean performance may have been due to the fact that there was absenteeism and considerable shifting of soldiers among crew positions (crew turbulence), especially during the early stages of training. As a result, the composition of crews (in particular, the TC/GNR combinations) varied somewhat from event to event. In that sense, the performance profiles were not traditional

learning curves wherein each sampling unit (here, a crew) serve as its own control. To examine within-crew changes in performance more closely, the data were first divided into two sets corresponding to training on the tank and training on GUARD FIST I. Then, every unique TC/GNR combination that participated in gunnery training was identified. The performance data from these crews were then reformatted to identify successive iterations (i.e., trials) of training sessions, regardless of intervening events.

These data are plotted as a function of trials on GUARD FIST I and on the tank in Figures 34 and 35, respectively. Contrary to expectations, the graph shows that performance does not increase as a function of trials. In fact, there appears to be a slight decrease in performance from trial 1 to trial 2 on GUARD FIST I; however, a repeated measures analysis of variance (ANOVA) indicated that none of the differences across trails was statistically reliable. Two hypotheses can be provided to explain the lack of a learning effect. First, the trials were sometimes separated by more than one month and/or by other training events. Forgetting may have nullified gains in learning. Although the hypothesis is quite reasonable, no direct evidence can be provided in the present data. Second, despite instructions to the contrary, the trainers may have been rating a more general impression of the crew's proficiency rather than performance during the training event. The crews had many years of experience in armor; thus, a few trials on a particular training method would have little or no effect on this overall judgment of gunnery proficiency. For whatever reason, the data failed to reveal a meaningful relationship between practice on the alternative training methods and gunnery proficiency.

Between-crew differences in training. Another important resource tradeoff is the relationship between amount of training and performance on the terminal criterion, Table VIII. To measure amount of training, a "session" was defined as a period of continuous training for a crew. Our observations indicated that the length of each of the session varied somewhat. GUARD FIST I sessions were approximately 20-60 mins long, whereas on-tank training sessions lasted as short as 15 mins to as long as several hours when mechanical difficulties were encountered. Unfortunately, the length of each session was not recorded on the Training Event Inventories. Despite these differences in time, most sessions consisted of approximately 5-20 gunnery engagements. Therefore, it was reasonable to operationalize amount of training as the number of separate training sessions in which a crew, a TC, or gunner performed their assigned roles.

To examine the correlation between amount of training and Table VIII performance, it is necessary to establish that there is sufficient variability among crews in amount of training. Table 6 presents data from crews who participated in Table VIII, presenting means and standard deviations for the numbers of sessions conducted on both GUARD FIST I and the tank. In addition to the number of training sessions for the crews, the number of training sessions that the TC and GNR practiced with those training methods were tabulated separately. This was done to account for the times when the TC and GNR were trained outside of the context of their own crews. The number of total training sessions was divided between those occurring prior to AT and those occurring during AT, because of differences in recency of training. In other words, the amount of training during AT (i.e., more recent training) should have had more effect on Table VIII than the training that had occurred months before AT. Note that the number of training sessions on the tank

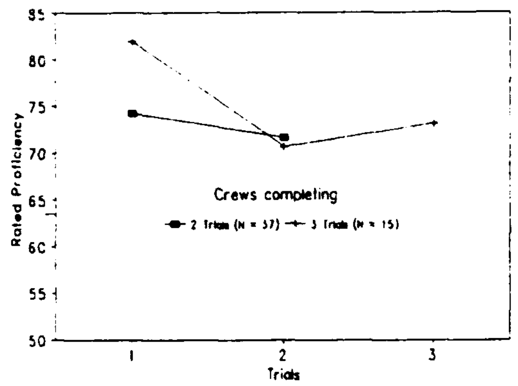

Figure 34. Rated proficiency as a function of trials on GUARD FIST I.

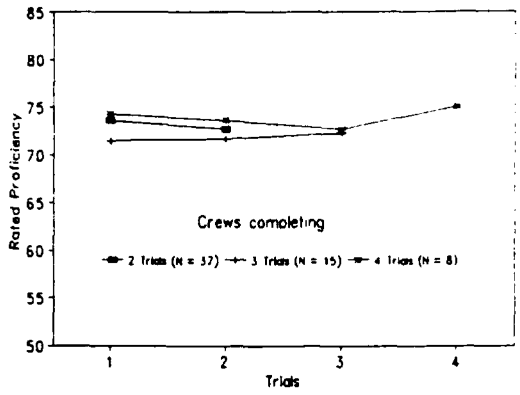

Figure 35. Rated proficiency as a function of trials on the tank.

during AT do *not* include the two live-fire sessions (day and night) that all crews performed for Table V/VI and Table VIII itself, because these numbers would be constants. These data indicate that crews averaged less than 2 sessions of GUARD FIST I and less than 3 sessions on the tank. (Note that Table V/VI and VIII would add an additional 4 on-tank sessions to that average.) Despite the relatively low number of training sessions, the standard deviations indicated that there was considerable variability among crews in those values.

Table 6

Means and Standard Deviations of Numbers of Training Sessions on GUARD FIST I and on the Tank

For Training	Before AT		During AT	
Conducted	M	(SD)	M	(SD)
On GUARD FIST I				
for the crew	0.4	(0.58)	0.8	(0.82)
for the TC	0.5	(0.67)	0.9	(0.84)
for the GNR	0.5	(0.65)	0.9	(0.81)
On the Tank				
for the crew	1.1	(1.29)	1.0	(0.64)
for the TC	1.9	(1.33)	1.0	(0.64)
for the GNR	1.6	(1.42)	1.1	(0.64)

The correlations between amount of training and performance on Table VIII are provided in Table 7. The correlations between amount of training with Table VIII were in fact stronger during AT than before AT. However, the correlations were in the negative direction: A greater number of training sessions was associated with poorer performance on Table VIII. The most likely explanation for this finding was provided by anecdotal reports from the unit trainers. During AT, trainers reported that their strategy was to ensure that weaker crews were targeted for extra training on GUARD FIST I and on the TCPCs.

Table 7

Correlations Between Number of Training Sessions and Table VIII Scores

For Training Conducted	Before AT	During AT
On GUARD FIST I		
for crew	.15 (41)	-.29 (41)
for TC	-.02 (41)	-.32* (41)
for GNR	.06 (41)	-.32* (41)
On the Tank		
for crew	.14 (32)	-.35* (41)
for TC	-.10 (32)	-.35* (41)
for GNR	.15 (32)	-.38* (41)

Note. Sample sizes for correlations are provided in parentheses.

*$p < .05$

Three inferences may be drawn about the negative relationships between amount of training and Table VIII scores. First, assuming the trainers adopt this compensatory training strategy described above, the negative correlations provide additional evidence that trainers can reliably assess crew gunnery proficiency. Second, to the extent the compensatory training strategy should reduce these correlations, the negative correlations provided no evidence about the effectiveness of either training intervention (GUARD FIST I or on-tank training). Third, the relation between amount of training and

performance on Table VIII is spurious; that is, the relation is caused by a third variable--the compensatory training strategy. Because of this spuriousness, the relation is not useful for establishing tradeoffs.

Comments on the Qualitative Data

Observations of the gunnery training events discussed in the previous section were made to help the interpretation of the data analyses. The observations are also valuable in their own right if used in the spirit of striving to continually improve the training process. The comments that follow concern issues that appear important to gunnery training for the Army in general and for the Reserve Component in particular. They should not be interpreted as criticisms. Certainly some of the comments stem from perceived weaknesses in training practices, but many are based on strengths that were observed. Other comments stem from variations in procedures. Thus, the comments are not pinpointed to particular deficits, but to a variety of areas that are mentioned because they are critical to the training process. Also, several of the comments concern issues that may not be under the control of the local command. These comments fall into five categories:

- training preparations,
- training procedures,
- MILES training,
- GUARD FIST I training, and
- other on-tank training

The domain of potential gunnery engagements is immense. Although the basic schema is constant -- find the target, manipulate the tank to fire on the target, assess the results, and continue -- there are numerous nuances concerning number and types of targets, condition of the tank, etc. which are combined in literally countless ways.[10] Furthermore, the procedures required to react to these conditions must be

- executed correctly,
- by a team of soldiers,
- with a coordinated effort,
- in a matter of a few seconds, and
- under life-threatening conditions.

The once-a-month training schedule for the Reserve Component compounds the training problem in two ways. First, the total amount of training time is severely constrained. Second, the long intervals between gunnery training events causes forgetting to be a significant factor. Thus, the basic thesis that underlies all of the following comments is that gunnery training time is at a premium. Any planning, coordination, or doubling up of efforts to expand the amount of productive time that can be devoted to practice should be pursued as relentlessly as possible. Passing Table VIII is a *minimum* criterion for successful gunnery. The ultimate goal for training should be far beyond that level.

---

[10]There have been several attempts to count the number of combinations with results that range from 64 (FM 17-12-1, DA 1988) to 4,618 (Meade, 1989).

Training Preparations

Every detail involved in gunnery performance becomes a detail to be concerned with in training. Moreover, training may be more complicated than performance in battle. Unlike combat, in which the enemy and the battlefield provide the targets and performance context, training must also be concerned about a variety of ancillary training details including targets, target operating equipment, ranges, monitoring devices (such as jump radios), and feedback devices (such as MILES), none of which would appear on the battlefield. Consequently, planning and advanced preparations are crucial. There are a number of items/issues that we would like to emphasize.

Detailed pre-training preparation is vital. To the maximum extent possible, training conditions should be set up prior to the weekend. In addition, there should be an organized assignment of last minute weekend setup duties according to soldiers' skills. For example, jump radio experts should install jump radios for on-tank exercises; the radios seem to be temperamental otherwise. This may require a concerted effort to train a sufficient number of soldiers to use training equipment so that three or four key unit soldiers are not trying to do all of the preparations.

Unit training is a boot-strapping process whereby unit leaders play a dual role: refreshing themselves at the same time they are teaching their crews. That puts a lot of pressure on platoon leaders, platoon sergeants, and TCs to be ready. On the other hand, time for them to prepare themselves for training is at a premium. Perhaps the use of easy-to-use, take-home materials, such as the ARI fire command study guide, "Engaging Targets with the M1/M1A1 Tank," (Drucker, 1991) would help.

A related leadership issue is the orchestration of soldiers during "down time." Concurrent training is usually planned for platoons to accomplish while not on the gunnery range or in the gunnery simulator. Despite these plans, there is also time created by unexpected, and often uncontrollable, delays. Waiting around to train is deadly to motivation. Making contingency plans for training during these periods could lead to an increase in productive training time. Platoon leaders should write training plans for how their soldiers should prioritize their time during the weekend, including how they might be able to conduct ad hoc training to get ready for their tank table runs or for their simulator time. For example, going over the wide variety of potential fire commands, subsequent fire commands, degraded fire commands, and all of their associated crew duties does not require any equipment. TCs may need some form of a guide to help them because they may have a tendency to underestimate the variety of fire commands available and to overestimate the ability of their crews to respond to them. Again there should be a relentless pursuit for continual improvement.

Unit leadership should spend time systematically reviewing their training processes. The assumption should be that, no matter how well training events were executed, they are sufficiently complicated that they can always be improved. The focus of the review should be on what happened this time and what could be done better next time rather than on how proficient crews are. Estimates of proficiency seemed to be overly optimistic which can make training too relaxed. Furthermore, proficiency cannot be improved without focusing on the training process. In other words, carefully attend to the training process and crew proficiency will take care of itself.

Instituting some of the Army's Total Quality Management (TQM) techniques may be helpful.

Finally, absences and unusual events can create havoc in training intact crews. Any efforts and plans that can minimize such disruptions will aid the training process.

Training Procedures

The comments in this section are concerned with training procedures in general and apply regardless of the method of training (e.g., MILES, GUARD FIST I, TCPC).

Because of the long intervals between training events, there is need for deliberate warm-up and review period before on-tank or simulator exercises. This period should involve review of fire commands, sight pictures, switchology, and other crew duties. If training is on MILES, alterations in gunnery procedures introduced by MILES should be discussed. Systematic review guidelines for this preparatory period should be developed. Such guidelines could also serve as training guidance for the ad hoc training advocated above.

TCs should review gunnery procedures while training is in progress. In addition to reviewing the execution of a gunnery engagement just completed, they should review procedures, as needed, to prepare their crews for up-coming exercises. For example, the Reserve Component Tank Gunnery Training Program, ST 17-12-RC (U.S. Army Armor School, 1989) and ARI Specifications for Tank Gunnery Table VIII, M1 (Drucker, 1991) provide instructional materials for reviewing crew duties for engagements based on cues that match the Table VIII scenarios. Thorough review prior to the training run plus coaching before each engagement should help make execution of training exercises less frustrating and more productive.

During tank tables administered by the companies, the tower controllers and assistant instructors (AIs) tend not to give feedback or assistance during exercises but wait for an after-action review (AAR). In general, when engagements differ such that mistakes in one are not likely to be repeated in subsequent engagements, the AAR is a appropriate mechanism for review. On the other hand, soldiers operating the tower should assist the TC with feedback during the exercise if the feedback might help them avoid repeating the same mistakes in following exercises.[11] For instance, consistent switchology mistakes that can be detected by soldiers operating the tower should be corrected during execution of the table. Prior to Table VIII, training is the primary objective, so there should be adaptive attention to the needs of each crew.

Soldiers operating the tower also need to warm-up and review to insure that they can quickly recognize incorrect procedures including gunnery procedures, MILES procedures, and other range procedures. Prepared scenarios should include notes concerning correct fire commands, switchology, and special MILES procedures. Gunnery exercises are not simple, and there always seem to be surprises in how procedures can go wrong.

---

[11]Note that this suggestion is congruent with guidelines for conducting AARs found in ST 17-12-RC (US Army Armor School, 1989).

In the same vein as anticipating and avoiding surprises, crews should double and triple check their equipment (e.g., radios, boresight, night sights, and MILES reset) to be absolutely certain they are ready to run a table. There seems to be an inverse relationship between confidence in being ready and actually being ready: the "no problem" response should be interpreted to mean the respondent has not thought about getting ready yet. Checklists, such as the training aids for preparing crew stations, may help, but the issue may be motivational. Again, a spirit of relentless pursuit of improvement may keep attention focused on the details needed to improve training of gunnery procedures.

Scheduling difficulties can create a reverse "hurry-up-and-wait" syndrome where crews must wait, then be hurried to get though the tank tables. Such scheduling creates pressure for getting all crews through training as opposed to getting performance as good as possible. Improving performance by repeating poorly performed exercises takes time. To the extent that fuel and ammunition allow, the goal of training should be to have crews repeat as many exercises as possible instead of just getting crews through an exercise at least once. This comment does not solve any of the time delay problems, but is offered only to reinforce efforts to minimize delays. A rule of thumb should be that everything takes three times as long as expected; consequently, there is no time to waste.

Given the probability of unanticipated delays, units may want to consider training less proficient crews first. This period is more appropriate for training poor crews because the soldiers operating the tower and AIs are fresh, the time crunch (e.g., approaching sunset) is not yet critical, and the crews can be given better attention and more repetitions.

One final comment is offered to stimulate more thorough analysis. What are realistic expectations for the Reserve Component's level of gunnery proficiency? The GUARD FIST I and M-COFT training matrices demand considerable training time (Morrison, Campshure, and Doyle, 1991). To become highly proficient at gunnery, gunnery training (and whatever else could fit in as concurrent training) could take up all of the available training time. Thus, the primary training allocation problem for the ARNG is not how many hours of what device to use, but how to pack more practice into the time available.

MILES

MILES equipment with target lifters were used for Tables I through IV, but use of MILES was avoided at AT because of its perceived negative transfer to live-fire gunnery. This raises a number of questions for MILES training: What skills are being trained, what skilled are not being trained, and what bad habits are being practiced? What is the plan following MILES to provide make-up practice for skills not trained on MILES and to extinguish bad habits acquired on MILES? MILES may provide gun manipulation practice but targets are generally stationary and research (e.g., Graham and Smith, 1990) has shown that this is not a difficult skill to learn. MILES can give hit/miss feedback (if everything is working), but that may only serve as motivational feedback because informational feedback from MILES is poor. Degraded mode engagements should probably not be practiced on MILES at all because the sight picture requirements for the GAS are wrong. Thus, the hit/miss feedback may even misplace the emphasis in training from practicing to gain speed using correct

gunnery procedures to practicing to hit targets with a wide-band laser using faulty switchology. Given the amount of preparation time needed to install MILES and the frequency with which it subsequently fails to operate causing training delays, one has to wonder why is MILES used at all? Dry-fire using full-up procedures, including ranging where permissible, may be as beneficial for Tables I through IV. With dry-fire, the tower cannot call target hits, but with the frequency that MILES fails, they often cannot anyway. Without MILES, trainers can focus their attention on making sure gunnery procedures are executed quickly and correctly.

<u>GUARD FIST I</u>

As argued earlier, GUARD FIST I has a tremendous advantage over on-tank training in terms of the time savings it can provide. It takes an incredibly long time for one crew to prepare for, execute, and receive a debriefing on one run of a Table. Table IV training, where one crew is trained at a time, means that only ten engagements (plus an occasional repeat for some crews) are fired per crew per weekend. In contrast, GUARD FIST I (or M-COFT for that matter) can give many more repetitions and cover more of the potential gunnery domain, particularly if two devices are available and used efficiently.

The observed ARNG battalion used GUARD FIST I training to augment, and in one case replace, FM 17-12-1 (DA, 1988) tank table training conducted on the tank. Because of scheduling considerations, however, the four companies varied in terms of where in the FM 17-12-1 sequence they were able to insert GUARD FIST I practice. Consequently prior to AT, one company conducted GUARD FIST I training using engagements that approximated the contents of Tank Table III and two others trained with engagements that approximated the contents of Table IV. At the beginning of AT, all companies trained on GUARD FIST I with a set engagements of exercises selected to approximate the contents of Table VII and VIII. Based on these experiences, several comments may be made about using GUARD FIST I most efficiently.

First, for GUARD FIST I training to be productive, crews must receive an orientation to the device. GUARD FIST I provides a familiarization sequence which should be used to acquaint crews with the graphics and with the relatively few idiosyncracies of operating the tank in GUARD FIST I mode (e.g., operation of buttons to simulate loading and tank speed indicated on the driver's monitor). All crews new to GUARD FIST I should receive this orientation. In addition, all crews, whether new to GUARD FIST I or not, should be allowed to perform several simple engagements to at the beginning of GUARD FIST I training to get the feel of the device. This will also serve as part of the warm-up and review advocated above for every training session. If the beginning of GUARD FIST I training is not conducted carefully, the crews may have difficulties in remembering and performing gunnery procedures. Incomplete preparation for GUARD FIST I practice can lead to frustration at not being able to perform well. This can result in negative consequences for motivation to learn.

Second, GUARD FIST I instructor/operators (I/Os) need to be actively involved in training. They must know gunnery procedures (fire commands, switchology) and they must be able to interpret GUARD FIST I error statements so that they can diagnose crews' deficiencies and give them assistance in improving. To assist the warm-up process, I/Os should give a warm-up talk on particular parts the gunnery to look out for. A list of common gunnery

Scanning techniques	Rules of lay:
3X to 10X searching	• G pattern
Dumping lead	• Center mass
TC verifying range	• Remember sight picture
Driving speed	Hear "Up" before issuing "Fire"
Most dangerous target first	Arming gun on battlecarry
Driver clearing berm at night	Acquisition reports
Reading GAS	Driver ID targets
Read back range of subsequent fire command	Gun select switch on "GUNNER"
	Lay and lase in hole on defense

Figure 36. List of gunnery procedures to emphasize in GUARD FIST I training compiled by an ARNG training NCO.

mistakes was compiled by one of the companies (see Figure 36). A "how-to-train with GUARD FIST I" manual should be prepared that includes procedures for warm-up and review at the beginning of each exercise and guidelines on giving coaching and feedback during training.

Third, while the battalion attempted to select exercises to match selected tank tables, the particular sequence of training is probably not as important as allowing crews to have sufficient time to use the GUARD FIST I training strategy to its advantage as recommended by Morrison, Campshure, and Doyle (1990). Results from the quantitative data suggested that one or two sessions on GUARD FIST I are not enough to appreciably affect performance. At the same time, the operation of GUARD FIST I appears simple enough to allow the I/O to select engagements from different parts of the GUARD FIST I program, so that GUARD FIST I practice can be congruent with a training schedule that includes GUARD FIST I with training on other devices and on the tank. However, it seems more important for the I/O to select exercises that are responsive to the needs of the crew.

Fourth, GUARD FIST I may be used advantageously throughout gunnery training. It may be particularly useful for concurrent training on weekends when on-tank training is also being conducted. For example, one company conducted on-tank training on Tables I, II, and III and also provided crews with approximately two hours of GUARD FIST I familiarization and training, all in one weekend. Densely packed weekends like this should be the training norm.

Finally, to fully take advantage of GUARD FIST I, setup should be completed prior to the weekend. Then training can proceed as soon as crews are present. That saves an even greater amount of time compared to typical on-tank training.

GUARD FIST I training is rigorous. The device does not allow short-cuts and is capable of detecting a wide range of errors that may escape notice during on-tank training. Furthermore, the engagements are more demanding than those that typically can be laid out on the range facilities most commonly available to weekend training. At the same time, GUARD FIST I training may provide a more realistic picture of the challenge of a Table VIII conducted on

a modern computer-controlled, Multi-Purpose Range Complex (MPRC). Thus, GUARD FIST I, if used properly, can provide excellent preparation for live-fire gunnery.

Live-Fire and Other On-Tank Training

Again, because training time is at a premium, efforts should be made to look for ways to increase the amount of productive time crews spend practicing on their tanks. This may mean loosening the strict lock-step schedule of the FM 17-12-1 to accommodate the variety of training methods available. For example, tank table training is typically conducted with one tank at a time.[12] If GUARD FIST I training can replace some of those tables, on-tank training time prior to live fire could be used for making adjustments to the feel of the controls of a "live tank." A formal tank table exercise may not be needed for that. Given a minimal amount of space, two or more tanks could simultaneous practice driving, target acquisition, and tracking skills in a TCPC-like format using the weekend time available to a much greater advantage.

Finally, the rigor and challenge of GUARD FIST I (or M-COFT) training should be matched during on-tank training. To match to the exactitude of the GUARD FIST I, soldiers operating the tower, AIs, and each TC must concentrate on the details of performance and not be too lenient in their review.

Conclusions

The results from the case study of gunnery can be summarized by a series of conclusions:

1. The correlations among proficiency ratings indicate that Army trainers are able to validly rate relative gunnery proficiency. At the same time, between-rater differences in strictness and leniency can cause calibration problems that make tracking performance across different events impossible.

2. The case study methods were not useful for determining resource-proficiency tradeoffs. The essentially negative finding for the nonexperimental method further bolsters the argument put forth in Chapter 2 that resource-proficiency relationships should be determined using experimental methods.

3. A specific objective of the research was to allocate training time to GUARD FIST I. The research did not result in a valid function describing the relationship between amount of training and proficiency; therefore, the allocation cannot be made on that basis. The negative results did, however, suggest that more than 1 or 2 sessions on the device is required to positively affect gunnery proficiency.

4. Training time is at a premium. GUARD FIST I (and M-COFT) have the ability to compact many more engagements per weekend then typical on-tank training. They should be used aggressively to increase the total amount of gunnery practice crews receive.

---

[12] Even when more than one tank is on the course, only one tank at a time is active.

5. The primary training allocation problem for the ARNG is not how many hours of what devices to train. The primary issue is how to pack more practice, by whatever training method, into the allotted time. This of course is not the training-tradeoff issue.

REFERENCES

Bessemer, D.W. (1990). Transfer of SIMNET training in the Armor Officer Basic Course (ARI Technical Report 920). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 198)

Bickley, W.R. (1980). Training device effectiveness: Formulation and evaluation of a methodology (ARI Research Report 1291). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A122 777)

Bliss, J.P. (1989). The APTS Battery as a surrogate measure for tank gunnery simulator performance. Unpublished master's thesis, University of Central Florida.

Boldovici, J.A. (1987). Measuring transfer in military settings. In S. Cormer, & J. Hagman, eds. Transfer of learning. New York: Academic Press.

Boldovici, J.A. (1986). [Description of an evaluation of VIGS]. Unpublished report.

Brennan, R.L. (1983). Elements of generalizability theory. Iowa City, IA: ACT Publications.

Campbell, C.H., & Campbell, R.C. (1990). Methodology for defining and sampling from the domain of threat conditions for crew and platoon tactical gunnery (ARI Technical Report 886). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A221 094)

Campbell, C.H., & Hoffman, R.G. (1990). Sampling the threat domain for efficient tank gunnery training and testing (ARI Research Report 1552). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A220 612)

Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally & Company.

Campshure, D.A. (1990). Devices and aids for training M1 tank gunnery in the National Guard: A detailed analysis of conditions and actions (ARI Research Report 1588). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A240 931)

Campshure, D.A., & Drucker, E.H. (1990). Predicting first-run gunnery performance on Tank Table VIII (ARI Research Report 1571). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A228 201)

Caro, P.W. (1977). Some factors influencing Air Force simulator training effectiveness (HumRRO Technical Report 77-2). Alexandria, VA: Human Resources Research Organization.

Carter, G., & Trollip, S. (1980). A constrained maximization extension to incremental transfer effectiveness or, how to mix your training technologies. Human Factors, 22, 141-152.

Clark, C., & Williges, R.C. (1973). Response surface methodology central-composite design modifications for human performance research. Human Factors, 15, 295-310.

Cook, T.D.; Campbell, D.T.; & Peracchio, L. (1990). Quasi-experimentation. In M.D. Dunnette, & L.M. Hough, eds. Handbook of industrial and organizational psychology: Vol. 1 (2nd ed., pp. 491-576). Palo Alto, CA: Consulting Psychologists Press.

Cronholm, J.N. (1985). The optimization of training systems. Proceedings of the Symposium on the Transfer of Training to Military Operational Systems (pp. 237-259). Brussels: North Atlantic Treaty Organization.

Daedalean, Inc. (1990). Guard Unit Armory Device, Full-Crew Interactive Simulation Training-Armor (GUARD FIST I) (TM9-6920-GFI-12-1, Change No. 1). Orlando, FL: Project Manager for Training Devices.

Department of the Army. (1988). Tank Combat Tables M1 (Field Manual 17-12-1). Washington, DC: Author.

Doyle, E.L. (1990). Threat presentations for selected battlefield scenarios (ARI Research Product 90-14). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A221 289)

Drucker, E.H. (1991). The development of a rapid trainup package and platoon-level scenarios for armor training the Army National Guard (HumRRO Final Report FR-PRD-91-11). Alexandria, VA: Human Resources Research Organization.

Graham, S.E., & Smith, T.L. (1991). Identifying tank gunnery skill requirements on the Institutional Conduct-of-Fire Trainer (I-COFT) (ARI Research Report 1583). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A235 597)

Hagin, W.V.; Osborne, S.R.; Hockenberger, R.L.; Smith, J.P.; & Gray, T.H. (1982). Operational test and evaluation handbook for aircrew training devices: Operational effectiveness evaluation (AFHRL Technical Report 81-44(II)). Brooks AFB, TX: Air Force Human Resources Laboratory.

Hammerton, M. (1989). Tracking. In D.H. Holding, ed. Human skills (2nd ed.). Chichester, England: John Wiley & Sons, Ltd.

Harris, J.H.; Ford, P.; Tufano, D.R.; & Wiggs, C. (1985). Application of transfer forecast methods to armor training devices (ARI Technical Report 669). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A172 210)

Hart, R.J.; Hagman, J.D.; & Bowne, D.S. (1990). Tank gunnery: Transfer of training from Top Gun to the Conduct-of-Fire Trainer (ARI Research Report 1560). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A223 165)

Hoffman, R.G. (1989). Description and prediction of Grafenwoehr M1 Table VIII performance (ARI Technical Report 837). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD B136 331)

Hoffman, R.G.; Fotouhi, C.H.; Meade, G.A.; & Blacksten, H.R. (1990). Design of threat-based gunnery performance test: Issues and procedures for crew and platoon tank gunnery (ARI Technical Report 893). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A226 356)

Hoffman, R.G., & Morrison, J.E. (1988). Requirements for a device-based training and testing program for M1 gunnery: Volume 1. Rationale and summary of results (ARI Technical Report 783). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A194 808)

Kraemer, R.E., & Bessemer, D.W. (1987). U.S. tank platoon training for the 1987 Canadian Army Trophy (CAT) competition using a Simulation Networking (SIMNET) system (ARI Research Report 1457). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A191 076)

Kraemer, R.E., & Smith, S.E. (1990). Soldier performance using a part-task gunnery device (TopGun) and its effects on Institutional Conduct-of-Fire trainer (I-COFT) proficiency (ARI Research Report 1570). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A227 403)

Lane, N.E. (1987). Skill acquisition rates and patterns: Issues and training implications. New York: Springer-Verlag, Inc.

Mazur & Hastie (1978). Leaning as accumulation: A reexamination of the learning curve. Psychological Bulletin, 85, 1256-1274.

McDaniel, Schmidt, and Hunter (1988). Job experience correlates of job performance. Journal of Applied Psychology, 73, 327-330.

Meade, G.A. (1989). M1 tank gunnery: A detailed analysis of conditions, behaviors, and processes (ARI Research Report 1532). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A217 416)

Morrison, J.E. (1990). Power analysis of gunnery performance measures: Differences between means of independent groups (ARI Technical Report 872). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A219 917)

Morrison, J.E.; Campshure, D.A.; & Doyle, E.L. (1991). A strategy for training and sustaining M1 tank gunnery skills in the Army National Guard (HumRRO Final Report FR-PRD-90-27). Alexandria, VA: Human Resources Research Organization.

Morrison, J.E.; Drucker, E.H.; & Campshure, D.A. (1990). Devices and aids for training M1 tank gunnery in the National Guard: A review of military documents and the research literature (ARI Research Report 1586). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A240 628)

Morrison, J.E., & Holding, D.H. (1990). Designing a gunnery training strategy (ARI Technical Report 899). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A226 129)

Morrison, J.E.; Meade, G.A.; & Campbell, R.C. (1990). Identification of tactical gunnery subtasks: Objectives for a threat-based training program (ARI Research Note 90-10). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A219 918)

Morrison, J.E., & Walker, S.A. (1990). Effects of mental practice on tank gunnery performance (ARI Technical Report 873). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A219 916)

Pfeiffer, M.G., & Horey, J.D. (1988). Analytic approaches to forecasting and evaluating training effectiveness (Technical Report 88-027). Orlando, FL: Naval Training Systems Center.

Postman, L. (1971). Transfer, interference and forgetting. In J.W. Kling, & L.A. Riggs, eds. Woodworth & Schlosberg's experimental psychology (3rd ed.). New York: Holt, Rinehart and Winston.

Povenmire, H.K., & Roscoe, S.N. (1973). Incremental transfer effectiveness of a ground-based general aviation trainer. Human Factors, 15, 534-542.

Roscoe, S.N. (1972). A little more on incremental transfer. Human Factors, 14, 363-364.

Roscoe, S.N. (1971). Incremental transfer effectiveness. Human Factors, 13, 561-567.

Roscoe, S.N., & Williges, R.C. (1980). Measurement of transfer of training. In S.N. Roscoe, ed. Aviation Psychology. Ames, IA: The Iowa State University Press.

Spears, W.D. (1985). Measurement of learning and transfer through curve fitting. Human Factors, 27, 251-266.

Sticha, P.J.; Schlager, M.; Buede, D.M.; Epstein, K.; & Blackstein, H.R. (1990). Designing skill-based training: Skill characterization and preliminary model requirements (ARI Technical Report 897). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A227 237)

Tufano, D., & Evans, R. (1982). The prediction of training device effectiveness: A review of Army models (ARI Technical Report 613). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A146 937)

Wilkinson, L. (1988). SYSTAT: The system for statistics. Evanston, IL: SYSTAT, Inc.

U.S. Army Armor School, Weapons Department (1990). Reserve component tank gunnery training program. Fort Knox, KY: Author.

Appendix A

Training Prediction Questionnaire

Name____________________

TRAINING PREDICTION QUESTIONNAIRE

The purpose of this questionnaire is to obtain your opinion about the effectiveness of three different methods of training crew gunnery. You will be asked to use your experience to estimate a crew's proficiency after it receives various amounts of RC unit training using each method. Your answers will help us generate a composite estimate of the training effectiveness of each method. Your opinions, for example, will help to provide answers to questions such as the following:

- What aspects of gunnery does GUARD FIST I train best?
- How efficient is GUARD FIST I training compared to Table VIII live-fire training?

Different training methods emphasize different aspects of gunnery. Because of this, the questionnaire focuses on the important components of tank gunnery rather than on overall gunnery proficiency. You already should be familiar with these components because they were used on the Training Event Inventory, the data form that you used this spring and summer to describe performance during gunnery training. These components are repeated below:

SEARCH
- Crew searches between and during engagements.
- Crew searches 360°, concentrating in tank's primary sector.
- Crew scans entire sectors/performs detailed searches.

REACTION DRILLS
- Crew's reactions are immediate.
- Crew returns fire on contact.
- Crew turns tank/turret per tactical situation.

NORMAL MODE FIRE COMMANDS AND REENGAGEMENT
- TC gives timely, clear fire commands.
- TC selects proper ammo, selects and sequences targets correctly.
- TC gives proper corrections, if required.
- Crew members give timely, correct verbal responses (crew duties).
- Crew reengages missed targets.

TRACKING AND SWITCHOLOGY
- TC/Gunner use proper ranging/lasing techniques
- TC/Gunner use proper tracking techniques
- TC/Gunner remember to DUMP LEAD after engagement
- TC/Gunner maintain proper switch settings throughout engagement

DEGRADED MODE AND SUBSEQUENT FIRE COMMANDS
- The crew isolates/corrects/ compensates for degraded conditions ASAP.
- TC gives timely, clear fire commands suitable to degraded condition.
- TC selects proper ammo, selects and sequences targets correctly.
- TC specifies battlesight when appropriate.
- TC gives accurate target descriptions.
- TC gives brief, effective direction element when required.
- TC gives proper corrections, if required.
- Crew members give timely, correct verbal responses (crew duties).
- TC/gunner use standard adjustments/subsequent fire commands per crew or adjacent tank observations.

SPOT REPORTS
- Crew accurately reports threat type, number, and action.
- Crew accurately reports location (within 200 meters).
- Crew accurately reports friendly actions.
- Crew transmits SPOT reports ASAP.

ACQUISITION REPORTS
- Crew transmits brief, timely reports.
- Crew gives accurate target descriptions and locations

GUNNERY COMPONENTS

In this questionnaire, you will be asked to estimate the proficiency of a novice tank crew after it receives different kinds of gunnery training and different amounts of each kind. To make your estimates, assume that the novice crew consists of

- a new TC with two years of experience as a gunner and two years of experience either as a loader or driver, but no experience as a TC
- a new gunner with two years of experience as a loader or driver, but no experience as a gunner
- a new driver just out of OSUT
- a new loader just out of OSUT

Assume training begins with this novice crew.

Make your estimates using a scale ranging from 0 to 100. The estimates go in the table on the last page of this questionnaire.

The following chart describes the overall gunnery proficiency of tank crews at different points along the scale. Notice that a novice tank crew would be expected to achieve an overall proficiency level of 40.

Proficiency Level	
100	Crews at this level of proficiency could easily score 1000 points on Table VIII, maximize their Tactical Table C, and be part of a platoon that maximizes its Tactical Table I and Platoon ARTEP. Crews trained to this level would devastate the enemy on the battle field.
90	
80	Crews at this level of proficiency could pass Table VIII with 800 to 850 points and complete their Tactical Tables with a passable number of errors. These crews would very likely kill targets and survive the battle.
70	
60	Crews at this level would fail Table VIII with 500 to 600 points and complete their Tactical Tables with an unacceptable number of errors. They may hit a target or two on the battlefield, but their chances of survival are uncertain.
50	
40	Crews at this level would perform as if they had a new TC, a new gunner, and a driver and loader who were recent OSUT graduates. Their contribution to mission success and chances of survival would be low.
30	
20	Crews at this level of proficiency would be a detriment on the battlefield. They would not hit targets and would probably not survive.
10	
0	Personnel in crews at this level have received no tank training.

The table on the last page of the questionnaire contains eight columns. The first six columns represent different amounts of training (none, 10 tasks, 20 tasks, etc.). Use these columns to estimate a novice crew's proficiency on each gunnery component (such as search, reaction drills, and tracking and switchology) after receiving each of the six different amounts of training as a crew.

The seventh column represents the maximum amount of training possible. In this column estimate how good a novice crew could become on each gunnery component if it were to receive an unlimited amount of training with the specified method.

The eighth column asks for your confidence in your proficiency estimates. You are asked to judge your confidence on a 5-point scale.

IMPORTANT: You will be asked to make proficiency estimates for three different training methods: MILES, GUARD FIST I, and Live Fire. When you make your estimates for one of these methods, assume it is the only training method that the novice crew is getting for its RC unit training. For example, when you estimate the proficiency achieved from live fire training, assume that the crew has not trained as a crew using MILES, dry fire, snake boards, or any other gunnery training method.

NOTE: Gunnery components that are obviously not trained by a particular method have been omitted from the table.

Training methods and tank gunnery components	What proficiency level (from 0 to 100) would you expect a novice crew to achieve after firing each of the following numbers of tasks in unit training? (Assume the tasks are similar to those in GUARD FIST I and the combat tables.)						What is the highest profic-iency possible using only this training method?	How confident are you of your estimates? 1-Very 2-Quite 3-Somewhat 4-Slightly 5-Not at all
	0 Tasks	10 Tasks	20 Tasks	30 Tasks	40 Tasks	50 Tasks		
MILES (e.g., Tables II, III, and IV) ONLY								
Search								
Acquisition Reports								
Normal mode fire commands and reengagement								
Degraded mode fire commands and subsequent commands								
Spot reports								
GUARD FIST I ONLY								
Search								
Acquisition reports								
Reaction drills								
Normal mode fire commands and reengagement								
Degraded mode fire commands and subsequent commands								
Tracking and switchology								
Spot reports								
Live fire (e.g. Table VII or VIII) ONLY								
Search								
Acquisition reports								
Normal mode fire commands and reengagement								
Degraded mode fire commands and subsequent commands								
Tracking and switchology								

THANK YOU FOR YOUR ASSISTANCE!

Appendix B

Training Event Inventory Instruction and Rating Form

Training Event Inventory Instruction Sheet

1. A Training Event Inventory is to be completed by the trainer for each crew during each gunnery training event.

2. The trainer may be the Readiness NCO, an I/O for one of the devices, a platoon sergeant, or a platoon leader.

3. A training event may be a practice period on GUARD FIST, U-COFT, M-COFT, TOPGUN, or VIGS; or it may be a crew's run on a TCPC or on one of the Tank Tables.

4. Fill in last name and first initial for the trainer/rater and all participating crew members. Indicate whether loaders and driver are regular members of the crew or hot-seated for this particular training event by circling "Reg." or "Hot." U-COFT/M-COFT, TOPGUN, and VIGS will not have loaders and drivers.

5. Fill in date and time of training. This is important for examining how the sequence of events affect training.

6. *Indicate the exercises* or tasks trained during the event using the training device exercise numbers or Tank Table task numbers. Write in the number of times each exercise or task was practiced.

7. Describe the crew's gunnery proficiency at the end of training using the rating scale provided. Please note that the top of the scale, 100% proficiency, means that the crew is essentially perfect, and that its gunnery just cannot be any better. There will be few crews at this level.

8. Attach (staple) copies of device printouts and/or Tank Table score sheets.

9. Please feel free to make copies of completed forms to use for tracking crews' progress.

10. Use the pre-addressed, stamped envelopes to return the rating forms. The envelopes have postage for up to *30 completed forms.*

11. Please direct any questions about the Inventory to Dr. R. Gene Hoffman, (502) 942-3232.

Training Event Inventory

Trainer:______________ Date:__________

Time:__________

TC:________ Gunner:________ Loader:________ (Reg. or Hot) Driver:________ (Reg. or Hot)

Company:____ Platoon:_______ Tank:_______

Training Event: (Check one)

___ GUARD FIST ___ U-COFT/M-COFT ___ VIGS
___ TOPGUN ___ TCPC with MILES ___ Table I
___ Table II ___ Table III ___ Table IV

Description of Event: For TCPCs and Tank Tables, describe the tasks included and the number of repetitions of each task. For GUARD FIST, U-COFT/M-COFT, VIGS, and TOPGUN, write the number of each exercise that was trained and the number of times each exercise was practiced.

Exercise/Task	No. of Repetitions	Exercise/Task	No. of Repetitions

Training Performance

Nine components of gunnery are identified below, each with bullets indicating important parts of the component. For each component, rate the proficiency of the crew at the *end of this training event.* To provide a basis for making your proficiency rating, compare the performance of the actual crew to the following hypothetical crews:

Proficiency	
100%	*Without any further training*, this crew could easily score 1000 points on Table VIII, max its Tactical Table C, and be part of a platoon that maxes its Tactical Table I and Platoon ARTEP.
95%	
90%	
85%	
80%	*Without any further training*, this crew could pass Table VIII with 800 to 850 points and complete its Tactical Tables with a passable number of errors.
75%	
70%	
65%	
60%	*Without further training*, this crew would fail Table VIII with 550 to 600 points and complete its Tactical Tables with an unacceptable number of errors.
55%	
50%	
Below 45%	This crew performed like recent OSUT students and a new TC that had never practiced together.

Use the numbers from 50 to 100 (e.g, 50, 55, 60, etc) that describe the proficiency of the crew at the end of this training session. Rate each component of gunnery trained during the event. For components that could not be trained during the event, write an NA. Comments about performance may be written on the right of the page.

_____SEARCH COMMENTS

- Crew searches between and during engagements.
- Crew searches 360°, concentrating in tank's primary sector.
- Crew scans entire sectors/performs detailed searches.

_____ACQUISITION REPORTS

- Crew transmits brief, timely reports.
- Crew gives accurate target descriptions and locations

(OVER)

COMMENTS

_____REACTION DRILLS
- Crews reactions are immediate.
- Crew returns fire on contact.
- Crew turns tank/turret per tactical situation.

_____CONTACT REPORTS
- Crew immediately reports contact.
- Crew accurately reports direction and target types.
- Crew transmits brief, clear contact reports.

_____MOVEMENT
- Crew uses covered and concealed route or smoke.
- Crew coordinates movement w/adjacent tanks.
- TC selects appropriate primary, alternate, and supplemental positions.
- Crew properly occupies/moves between hide, turret down, and hull down positions.
- TC directs movement out of position to avoid AT fires.
- Driver maintains a steady firing platform and suitable speed.
- Crew avoids untrafficable terrain.

_____NORMAL MODE FIRE COMMANDS AND REENGAGEMENT
- TC gives timely, clear fire commands.
- TC selects proper ammo, selects and sequences targets correctly.
- TC gives proper corrections, if required.
- Crew members give timely, correct verbal responses (crew duties).
- Crew reengages missed targets.

_____DEGRADED MODE AND SUBSEQUENT FIRE COMMANDS
- The crew isolates/corrects/compensates for degraded conditions ASAP.
- TC gives timely, clear fire commands suitable to degraded condition.
- TC selects proper ammo, selects and sequences targets correctly.
- TC specifies battlesight when appropriate.
- TC gives accurate target descriptions.
- TC gives brief, effective direction element when required.
- TC gives proper corrections, if required.
- Crew members give timely, correct verbal responses (crew duties).
- TC/gunner use standard adjustments/subsequent fire commands per crew or adjacent tank observations.

_____TRACKING AND SWITCHOLOGY
- TC/Gunner use proper ranging/lasing techniques
- TC/Gunner use proper tracking techniques
- TC/Gunner remembers to DUMP LEAD after engagement
- TC/Gunner maintains proper switch settings throughout engagement

_____SPOT REPORTS
- Crew accurately reports threat type, number, and action.
- Crew accurately reports location (within 200 meters).
- Crew accurately reports friendly actions.
- Crew transmits SPOT reports ASAP.

920514